

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:31:47 ; Search time 97.4247 Seconds  
(without alignments)  
65.702 Million cell updates/sec

Title: US-10-645-659A-10

Perfect score: 86

Sequence: 1 TWHYYLNGRTATR 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

A\_Geneseq\_8:\*

1: Geneseq1990s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

9: Geneseq2005s:\*

10: Geneseq2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	86	100.0	14	8	ADR88216 Human hep
2	86	100.0	14	8	ADT78183 Functiona
3	86	100.0	14	9	AEA42432 Human hep
4	86	100.0	15	9	ADU71059 Human hep
5	86	100.0	15	9	ADU71060 Human hep
6	86	100.0	15	9	ADY27047 Heparanas
7	86	100.0	386	8	ADR88207 Human mat
8	86	100.0	386	8	ADT78174 45KDa sub
9	86	100.0	386	9	ADY27057 Heparanas
10	86	100.0	386	9	ADZ18995 Human hep
11	86	100.0	386	9	AEA42423 Human mat
12	86	100.0	460	9	ADY27061 Heparanas
13	86	100.0	486	9	AE887589 Human hep
14	86	100.0	492	9	ADZ18996 Hep106 co
15	86	100.0	493	9	AE887562 Human hep
16	86	100.0	495	9	ADZ18999 Hep109 co
17	86	100.0	497	9	AE887587 Human hep
18	86	100.0	501	9	ADZ19000 HepGS3 co
19	86	100.0	507	9	ADZ19005 HepGS6 co
20	86	100.0	508	9	ADY27058 Human ina
21	86	100.0	526	9	ADZ19006 HepHyalur
22	86	100.0	527	5	ABB07815 Chicken s
23	86	100.0	527	7	ABW02018 Chimeric

24	86	100.0	527	8	ADO63825	ADO63825 Chimeric
25	86	100.0	527	8	ADO63827	ADO63827 Chimeric
26	86	100.0	527	8	ADO63826	ADO63826 Chimeric
27	86	100.0	527	9	ADZ19004	ADZ19004 HepGS4 co
28	86	100.0	532	2	AA170083	AA170083 Seq ID No
29	86	100.0	543	2	AA170083	AA170083 Seq ID No
30	86	100.0	543	2	AA170082	AA170082 Human hep
31	86	100.0	543	3	AA170082	AA170082 Human hep
32	86	100.0	543	3	AA170082	AA170082 Human hep
33	86	100.0	543	3	AA170082	AA170082 Human hep
34	86	100.0	543	4	AA170082	AA170082 Human hep
35	86	100.0	543	4	AA170082	AA170082 Human hep
36	86	100.0	543	4	AA170082	AA170082 Human hep
37	86	100.0	543	5	AA170082	AA170082 Human hep
38	86	100.0	543	7	AA170082	AA170082 Human hep
39	86	100.0	543	7	AA170082	AA170082 Human hep
40	86	100.0	543	8	AA170082	AA170082 Human hep
41	86	100.0	543	8	AA170082	AA170082 Human hep
42	86	100.0	543	8	AA170082	AA170082 Human hep
43	86	100.0	543	8	AA170082	AA170082 Human hep
44	86	100.0	543	8	AA170082	AA170082 Human hep
45	86	100.0	543	8	AA170082	AA170082 Human hep
46	86	100.0	543	8	AA170082	AA170082 Human hep
47	86	100.0	543	8	AA170082	AA170082 Human hep
48	86	100.0	543	8	AA170082	AA170082 Human hep
49	86	100.0	543	8	AA170082	AA170082 Human hep
50	86	100.0	543	8	AA170082	AA170082 Human hep
51	86	100.0	543	8	AA170082	AA170082 Human hep
52	86	100.0	543	8	AA170082	AA170082 Human hep
53	86	100.0	543	8	AA170082	AA170082 Human hep
54	86	100.0	543	8	AA170082	AA170082 Human hep
55	86	100.0	543	9	AA170082	AA170082 Human hep
56	86	100.0	543	9	AA170082	AA170082 Human hep
57	86	100.0	543	9	AA170082	AA170082 Human hep
58	86	100.0	543	9	AA170082	AA170082 Human hep
59	86	100.0	543	10	AA170082	AA170082 Human hep
60	86	100.0	543	6	AA170082	AA170082 Human hep
61	86	100.0	545	7	AA170082	AA170082 Human hep
62	86	100.0	545	8	AA170082	AA170082 Human hep
63	86	100.0	545	8	AA170082	AA170082 Human hep
64	86	100.0	545	8	AA170082	AA170082 Human hep
65	86	100.0	545	8	AA170082	AA170082 Human hep
66	86	100.0	545	8	AA170082	AA170082 Human hep
67	86	100.0	545	8	AA170082	AA170082 Human hep
68	86	100.0	545	8	AA170082	AA170082 Human hep
69	86	100.0	545	8	AA170082	AA170082 Human hep
70	86	100.0	545	8	AA170082	AA170082 Human hep
71	86	100.0	545	8	AA170082	AA170082 Human hep
72	86	100.0	545	8	AA170082	AA170082 Human hep
73	86	100.0	545	8	AA170082	AA170082 Human hep
74	86	100.0	545	8	AA170082	AA170082 Human hep
75	86	100.0	545	8	AA170082	AA170082 Human hep
76	86	100.0	545	8	AA170082	AA170082 Human hep
77	86	100.0	545	8	AA170082	AA170082 Human hep
78	86	100.0	545	8	AA170082	AA170082 Human hep
79	86	100.0	545	8	AA170082	AA170082 Human hep
80	86	100.0	545	8	AA170082	AA170082 Human hep
81	86	100.0	545	8	AA170082	AA170082 Human hep
82	86	100.0	545	8	AA170082	AA170082 Human hep
83	86	100.0	545	8	AA170082	AA170082 Human hep
84	86	100.0	545	8	AA170082	AA170082 Human hep
85	86	100.0	545	8	AA170082	AA170082 Human hep
86	86	100.0	545	8	AA170082	AA170082 Human hep
87	86	100.0	545	8	AA170082	AA170082 Human hep
88	86	100.0	545	8	AA170082	AA170082 Human hep
89	86	100.0	545	8	AA170082	AA170082 Human hep
90	86	100.0	545	8	AA170082	AA170082 Human hep
91	86	100.0	545	8	AA170082	AA170082 Human hep
92	86	100.0	545	8	AA170082	AA170082 Human hep
93	86	100.0	545	8	AA170082	AA170082 Human hep
94	86	100.0	545	8	AA170082	AA170082 Human hep
95	86	100.0	545	8	AA170082	AA170082 Human hep
96	86	100.0	545	8	AA170082	AA170082 Human hep

97	74	86.0	545	9	ADY27034	Ady27034 Bovine he	170	43	50.0	981	5	ABB78727	Abb78727 Human cal
98	72	83.7	15	9	ADU70969	Adu70969 Human hep	171	43	50.0	981	7	ADE40182	Ade40182 Human NOV
99	70	81.4	15	9	ADU70927	Adu70927 Human hep	172	43	50.0	981	9	ADY70315	Ady70315 Human bet
100	67	77.9	15	9	ADU71058	Adu71058 Human hep	173	43	50.0	981	9	ADY70649	Ady70649 Human BAC
101	62	72.1	9	9	ADU70797	Adu70797 Human hep	174	43	50.0	1004	4	ABB65405	Abb65405 Drosophil
102	62	72.1	9	9	ADU70662	Adu70662 Human hep	175	43	50.0	1027	6	ABP75726	Abp75726 Human sec
103	62	72.1	15	9	ADU71272	Adu71272 Human hep	176	43	50.0	1177	8	AAB96721	Aab96721 Putative
104	62	72.1	15	9	ADU71210	Adu71210 Human hep	177	43	50.0	1177	8	ADS43128	Ads43128 Bacterial
105	56	65.1	9	9	ADU70700	Adu70700 Human hep	178	43	50.0	1200	8	ADN20018	Adn20018 Bacterial
106	56	65.1	9	9	ADU70638	Adu70638 Human hep	179	42.5	49.4	1035	7	ADC00940	Adc00940 Enterohae
107	56	65.1	15	9	ADU70968	Adu70968 Human hep	180	42.5	49.4	1035	9	AED81975	Aed81975 Hyperimmu
108	56	65.1	50	5	AAM50385	Aam50385 Mouse hep	181	42	48.8	56	4	AAU63539	Aau63539 Propionib
109	54	62.8	15	9	ADU71211	Adu71211 Human hep	182	42	48.8	56	6	ABM60058	Abm60058 Propionib
110	52	60.5	9	9	ADU70639	Adu70639 Human hep	183	42	48.8	218	6	ABU21887	Abu21887 Protein e
111	51	59.3	1179	9	ADS41964	Ads41964 Bacterial	184	42	48.8	223	4	AAU64792	Aau64792 Propionib
112	51	59.3	1179	8	ADN18677	Adn18677 Bacterial	185	42	48.8	223	6	ABM61311	Abm61311 Propionib
113	50	58.1	9	9	ADU70390	Adu70390 Human hep	186	42	48.8	297	4	ABBS2977	Abbs2977 Escherich
114	50	58.1	15	9	ADU70967	Adu70967 Human hep	187	42	48.8	322	8	ADX67425	Adx67425 Plant ful
115	49	57.0	9	9	ADU70516	Adu70516 Human hep	188	42	48.8	337	6	ABUS2328	Abu52328 Human GPC
116	47	54.7	9	9	ADU70342	Adu70342 Human hep	189	42	48.8	337	8	ADL23991	Adl23991 Human NOV
117	47	54.7	15	9	ADU71061	Adu71061 Human hep	190	42	48.8	337	8	ABO84700	Abo84700 Mouse can
118	46	53.5	15	9	ADU71209	Adu71209 Human hep	191	42	48.8	340	2	AAR65244	Aar65244 Human SF3
119	46	53.5	179	6	ABR40721	Abr40721 Momordica	192	42	48.8	340	7	ADD75541	Add75541 Prostata
120	46	53.5	179	7	ADN23627	Adn23627 Polypepti	193	42	48.8	340	7	ADD14165	Add14165 Human src
121	46	53.5	245	8	ADN24954	Adn24954 Bacterial	194	42	48.8	340	8	ABO84704	Abo84704 Human can
122	46	53.5	245	8	ADN22195	Adn22195 Bacterial	195	42	48.8	340	8	ABO84703	Abo84703 Human can
123	46	53.5	314	2	AAR75200	Aar75200 Rat P-F4M	196	42	48.8	340	8	ABO84702	Abo84702 Human can
124	46	53.5	350	2	AAR75198	Aar75198 Rat Gal-b	197	42	48.8	340	8	ABO84701	Abo84701 Human can
125	46	53.5	350	7	ADE59126	Ade59126 Rat Prote	198	42	48.8	343	2	AAR41670	Aar41670 Porcine s
126	46	53.5	350	7	ADE59130	Ade59130 Rat Prote	199	42	48.8	343	2	AAR65240	Aar65240 Porcine s
127	46	53.5	350	7	ADE64099	Ade64099 Human Pro	200	42	48.8	343	9	AED08895	Aed08895 Amino aci
128	46	53.5	350	7	ADE64097	Ade64097 Rat Prote	201	42	48.8	358	6	ABM64563	Abm64563 Propionib
129	46	53.5	350	7	ADDA45580	Add45580 Rat Prote	202	42	48.8	358	8	ADN60340	Adn60340 B. lichen
130	46	53.5	350	7	ABM85655	Abm85655 Human Pro	203	42	48.8	375	8	ADN17709	Adn17709 Bacterial
131	46	53.5	351	7	ADE59132	Ade59132 Human Pro	204	42	48.8	399	4	ABG18000	Abg18000 Novel hum
132	46	53.5	351	7	ADDA45582	Add45582 Human Pro	205	42	48.8	554	7	ADE56932	Ade56932 Rat Prote
133	46	53.5	351	7	ADE59128	Ade59128 Human Pro	206	42	48.8	554	7	ADE56928	Ade56928 Rat Prote
134	46	53.5	379	7	ABM85654	Abm85654 Mouse pro	207	42	48.8	574	2	AAR66033	Aar66033 Human ARD
135	44	51.2	2835	5	ABY98574	Abby98574 Dextran s	208	42	48.8	574	2	AAR66034	Aar66034 Rat ARD 1
136	44	51.2	2835	6	ABR55594	Abr55594 Amino aci	209	42	48.8	574	7	ADE56930	Ade56930 Human Pro
137	44	51.2	3972	3	AAB23749	Aab23749 S. avermi	210	42	48.8	574	7	ADE56934	Ade56934 Human Pro
138	44	51.2	3972	4	AAG65264	Aag65264 Streptomy	211	42	48.8	664	4	AAB93120	Aab93120 Human pro
139	44	51.2	3972	4	AAG65268	Aag65268 Streptomy	212	42	48.8	664	5	ABJ10928	Abj10928 Protein o
140	44	51.2	5532	3	AAB23752	Aab23752 S. avermi	213	42	48.8	664	5	AAU77036	Aau77036 Human TRC
141	44	51.2	5532	4	AAG65267	Aag65267 Streptomy	214	42	48.8	664	6	ABJ37044	Abj37044 Human bre
142	43	50.0	181	4	AAB59819	Aab59819 Tute prot	215	42	48.8	664	7	ADB80932	Adb80932 RING-SH c
143	43	50.0	331	5	AAM50383	Aam50383 Human hep	216	42	48.8	664	8	ADR89544	Adr89544 Apoptosis
144	43	50.0	437	9	ABM90536	Abm90536 M. xanthu	217	42	48.8	664	8	ADX06417	Adx06417 Cyclin-de
145	43	50.0	439	4	AAU07423	Aau07423 Human hep	218	42	48.8	668	8	ADR89542	Adr89542 Apoptosis
146	43	50.0	470	5	AAE18328	Aae18328 Human hep	219	42	48.8	691	9	AED08889	Aed08889 Amino aci
147	43	50.0	480	4	AAI97634	Aay97634 Human hep	220	42	48.8	841	9	AED08890	Aed08890 Amino aci
148	43	50.0	480	4	AAU07418	Aau07418 Novel hum	221	42	48.8	956	5	ABY78729	Abby78729 Human cal
149	43	50.0	480	4	AAB85217	Aab85217 Heparanas	222	42	48.8	956	7	ADY70317	Ady70317 Human bet
150	43	50.0	492	4	AAB84664	Aab84664 Amino aci	223	42	48.8	968	7	ADC65993	Adc65993 Human XB3
151	43	50.0	528	5	AAE18327	Aae18327 Human hep	224	42	48.8	968	9	ADZ25519	Adz25519 Cytokine
152	43	50.0	534	4	ABB85216	Aab85216 Heparanas	225	42	48.8	968	9	AEA15828	Aea15828 Human alc
153	43	50.0	534	5	ABP69310	Abp69310 Human pol	226	42	48.8	968	9	AEC21698	Aec21698 Alkadiene
154	43	50.0	534	5	AAM50337	Aam50337 Human pre	227	42	48.8	1022	9	ADZ25508	Adz25508 Cytokine
155	43	50.0	538	4	AAI97633	Aay97633 Human hep	228	42	48.8	1144	9	AED09961	Aed09961 Pathogeni
156	43	50.0	581	8	ADU25303	Adu25303 Bacillus	229	42	48.8	1175	9	AED82435	Aed82435 Hyperimmu
157	43	50.0	582	5	AAE18327	Aae18327 Human hep	230	41	47.7	120	6	ABU31307	Abu31307 Protein e
158	43	50.0	592	4	AAI97632	Aay97632 Human hep	231	41	47.7	133	4	AAU54261	Aau54261 Propionib
159	43	50.0	592	4	AAB81062	Aab81062 Human hep	232	41	47.7	133	6	ABM50780	Abm50780 Propionib
160	43	50.0	592	4	AAU07424	Aau07424 Human hep	233	41	47.7	179	4	AAU18319	Aau18319 Human end
161	43	50.0	592	4	ABB85215	Aab85215 Heparanas	234	41	47.7	202	7	ABM89756	Abm89756 Rice abio
162	43	50.0	608	6	ABU20989	Abu20989 Protein e	235	41	47.7	226	8	ADN99749	Adn99749 Novel hum
163	43	50.0	953	7	ADZ28107	Adz28107 Human NTR	236	41	47.7	240	8	ADX87351	Adx87351 Plant ful
164	43	50.0	962	7	ADE28106	Ade28106 Human NTR	237	41	47.7	311	8	ADT58484	Adt58484 Plant pol
165	43	50.0	971	7	ADC65991	Adc65991 Human XB3	238	41	47.7	311	6	ABU32312	Abu32312 Protein e
166	43	50.0	971	9	ADY70619	Ady70619 Human BAC	239	41	47.7	428	2	AAR44403	Aar44403 Canine co
167	43	50.0	971	9	AEA15827	Aea15827 Human alc	240	41	47.7	435	7	ABO67223	Abo67223 Klebsiell
168	43	50.0	971	9	AEB06096	Aeb06096 Amino aci	241	41	47.7	479	6	ABP79127	Abp79127 N. gonorr
169	43	50.0	971	9	AEC21696	Aec21696 Alkadiene	242	41	47.7	490	5	ABP65335	Abp65335 Bifidobac



243	41	47.7	507	8	ADK16983	Nanoarcha	316	40	46.5	572	5	ABB47327	Listeria
244	41	47.7	511	4	AAU38104	Streptoco	317	40	46.5	646	4	ABG06358	Novel hum
245	41	47.7	511	9	ADY40315	iPGM SCO.	318	40	46.5	670	8	ADO61975	Transcrip
246	41	47.7	578	8	ADQ65875	Novel hum	319	40	46.5	720	8	ADN23654	Bacterioc
247	41	47.7	609	8	ADQ66688	Novel hum	320	40	46.5	1191	8	ADN46644	Thermoco
248	41	47.7	610	4	AAAB31664	Amino aci	321	40	46.5	2054	4	ABG23323	Novel hum
249	41	47.7	748	2	AAAR24396	Prod. of	322	40	46.5	2273	4	ABG23304	Novel hum
250	41	47.7	748	2	AAAR24398	Prod. of	323	40	46.5	5909	4	ABG233295	Novel hum
251	41	47.7	748	2	AAAR24513	Consensus	324	40	46.5	6619	4	ABG23329	Novel hum
252	41	47.7	749	2	AAAR24466	Feline in	325	39	45.3	9	ADU70763	Human hep	
253	41	47.7	749	2	AAAR24465	Feline in	326	39	45.3	15	ADU71057	Human hep	
254	41	47.7	819	8	ADP99142	Human tra	327	39	45.3	38	AAU17157	Peptide #	
255	41	47.7	908	3	AAAB42511	Human ORF	328	39	45.3	38	ABM36158	Peptide #	
256	41	47.7	1101	2	AAAR44401	Canine co	329	39	45.3	38	AAU29648	Peptide #	
257	41	47.7	1225	8	ADU08938	Coronavir	330	39	45.3	38	ABB30966	Peptide #	
258	41	47.7	1307	6	AAO31015	Human tra	331	39	45.3	38	ABB21544	Protein #	
259	41	47.7	1353	6	AAE29913	Human tra	332	39	45.3	38	AAU69327	Human bon	
260	41	47.7	1359	8	ADSF75981	Human ATP	333	39	45.3	38	ABG51002	Human liv	
261	41	47.7	1391	8	ADP99172	Human tra	334	39	45.3	38	AAU04851	Peptide #	
262	41	47.7	1433	2	AAAR27818	CCV-6 spi	335	39	45.3	38	ABG38943	Human pep	
263	41	47.7	1447	8	ABY00038	Transmiss	336	39	45.3	64	ABO65501	Klebsiell	
264	41	47.7	1449	8	ADN37292	PTGV spik	337	39	45.3	85	ABG30007	Novel hum	
265	41	47.7	1451	2	AAAR27819	CCVinsavc	338	39	45.3	90	AAE10718	Chicken 4	
266	41	47.7	1452	2	AAAR24271	Canine co	339	39	45.3	90	AAE10717	Pig 4ST3G	
267	41	47.7	1452	2	AAAR44400	Canine co	340	39	45.3	90	AAE10713	Rat 4ST3G	
268	41	47.7	1452	6	ABP98183	Amino aci	341	39	45.3	90	AAE10716	Mouse 4ST	
269	41	47.7	1453	2	AAAR27820	CCV-CS4 s	342	39	45.3	90	AAE10715	Human 4ST	
270	41	47.7	1453	8	AAU08932	E2 protei	343	39	45.3	90	AAE10714	Human ORF	
271	41	47.7	1453	8	ADU08932	Coronavir	344	39	45.3	122	3	AAU41865	Human ORF
272	41	47.7	1454	2	AAAR24511	Prod. of	345	39	45.3	133	4	AAU66016	Propionib
273	41	47.7	1454	2	AAAR24397	Prod. of	346	39	45.3	133	6	ABM62535	Propionib
274	41	47.7	1454	2	AAAR24395	Prod. of	347	39	45.3	149	7	ABO83324	Pseudomon
275	41	47.7	1454	2	AAAR24468	Feline en	348	39	45.3	173	4	AAU92528	Human dig
276	41	47.7	1454	2	AAAR24277	FECV/PIPV	349	39	45.3	173	4	AAU22565	Novel hum
277	41	47.7	1454	2	AAAR24273	FIPV/PECV	350	39	45.3	173	7	ADB32405	Human nov
278	41	47.7	1454	2	AAAR24274	FECV/PIPV	351	39	45.3	186	4	ABU53119	Intracell
279	41	47.7	1454	2	AAAR24272	FECV/PIPV	352	39	45.3	190	8	ADR10089	Human pro
280	41	47.7	1454	2	AAAR24267	Feline in	353	39	45.3	230	8	ADQ82666	Wild type
281	41	47.7	1454	2	AAAR24270	Feline in	354	39	45.3	240	4	ABM65939	Drosophil
282	41	47.7	1454	2	AAAR24275	FECV/PIPV	355	39	45.3	273	3	AAU27984	Arabidops
283	41	47.7	1454	2	AAAR24278	FECV/PIPV	356	39	45.3	277	3	AAU27983	Arabidops
284	41	47.7	1454	2	AAAR24264	Feline in	357	39	45.3	284	3	AAU07300	Arabidops
285	41	47.7	1483	1	AAAP80474	Sequence	358	39	45.3	293	8	ADJ48984	Oil-asso
286	41	47.7	1594	1	AAAP81183	Sequence	359	39	45.3	294	7	ABO43093	Oil-thalla
287	41	47.7	6239	3	AAAB23750	S. avermi	360	39	45.3	294	7	ADB31859	Plant (A.
288	41	47.7	6239	9	AAAG65265	Streptomy	361	39	45.3	294	7	ADU55772	Thalecres
289	40	46.5	15	9	ADU07098	Human hep	362	39	45.3	294	8	ADU02279	Thalecres
290	40	46.5	27	9	ADU71212	Modified	363	39	45.3	295	3	AAU07299	Arabidops
291	40	46.5	47	6	ABP77982	N. gonorr	364	39	45.3	305	4	ABG13155	Novel hum
292	40	46.5	66	3	ABP77982	N. gonorr	365	39	45.3	306	4	ABG08885	Novel hum
293	40	46.5	66	3	AAU59580	Arabidops	366	39	45.3	314	4	ABG13158	Novel hum
294	40	46.5	66	3	AAU59580	Arabidops	367	39	45.3	320	6	ABU37049	Protein e
295	40	46.5	74	5	AAU73012	Neisseria	368	39	45.3	321	6	ABU27982	Protein e
296	40	46.5	81	6	ABP77980	N. gonorr	369	39	45.3	326	6	ABU25302	Pucative
297	40	46.5	115	7	ADP08017	Bacterial	370	39	45.3	365	4	AAU96750	Plant pol
298	40	46.5	144	6	ABJ19391	Human int	371	39	45.3	389	8	ADT60416	Plant lip
299	40	46.5	244	5	ABB90262	Human pol	372	39	45.3	389	9	ABE14376	Plant lip
300	40	46.5	328	6	ABM68047	Phototrab	373	39	45.3	426	8	ADL04592	M. catarr
301	40	46.5	339	3	AAU51208	Arabidops	374	39	45.3	451	5	ABB92483	Herbicida
302	40	46.5	353	3	AAU51207	Arabidops	375	39	45.3	452	5	ABG93908	Herbicida
303	40	46.5	374	8	ADN25772	Bacterial	376	39	45.3	478	4	ABG08884	Novel hum
304	40	46.5	375	8	ADS42873	Bacterial	377	39	45.3	478	4	ABG13156	Novel hum
305	40	46.5	389	6	ABM68309	Phototrab	378	39	45.3	529	2	AAU82586	Human pat
306	40	46.5	390	6	ABU22412	Protein e	379	39	45.3	551	6	ABU28836	Protein e
307	40	46.5	396	8	ADG66961	Plant ful	380	39	45.3	558	6	ADA33583	Acinetoba
308	40	46.5	397	3	AAU51206	Arabidops	381	39	45.3	572	6	AAE33683	Human str
309	40	46.5	400	6	ABU23209	Protein e	382	39	45.3	576	4	ABE61241	Drosophil
310	40	46.5	464	8	ADU36976	Plant ful	383	39	45.3	583	4	ABM63650	Drosophil
311	40	46.5	508	7	ADM03819	Human pro	384	39	45.3	690	4	ABM63232	Drosophil
312	40	46.5	508	9	AEC86749	Human cdn	385	39	45.3	809	9	ADX05059	Horse IL4
313	40	46.5	509	4	ABG23315	Novel hum	386	39	45.3	809	9	ADX05094	Horse IL4
314	40	46.5	534	9	ADW95051	Lycopersi	387	39	45.3	809	9	ADX05095	Horse IL4
315	40	46.5	534	9	ADX05253	Lycopersi	388	39	45.3	809	9	ADX05093	Horse IL4

389	39	45.3	809	9	ADX05046	Adx05046	Horse IL4	462	38	44.2	256	2	AAy08564	Aay08564	B. subtili
390	39	45.3	822	4	ABG24362	Abg24362	Novel hum	463	38	44.2	260	8	ADx93633	Adx93633	Plant ful
391	39	45.3	822	4	ABG29163	Abg29163	Novel hum	464	38	44.2	275	6	ADA36243	Ada36243	Acinetoba
392	39	45.3	822	4	ABG08883	Abg08883	Novel hum	465	38	44.2	276	8	AO80530	Aob80530	Burkholde
393	39	45.3	825	7	ABO68417	AbO68417	Pseudomon	466	38	44.2	285	4	AAB95183	Aab95183	Human pro
394	39	45.3	861	4	ABG22537	Abg22537	Novel hum	467	38	44.2	285	7	ADC31422	Adc31422	Human nov
395	39	45.3	861	4	ABG18082	Abg18082	Novel hum	468	38	44.2	289	8	ADN73295	Adn73295	Thale cre
396	39	45.3	887	6	ABU45334	Abu45334	Protein e	469	38	44.2	289	8	ADx75184	Adx75184	Plant ful
397	39	45.3	889	6	ABU46941	Abu46941	Protein e	470	38	44.2	301	8	ADY09051	Ady09051	Plant ful
398	39	45.3	889	6	ABU47500	Abu47500	Protein e	471	38	44.2	302	6	ABU07872	Abu07872	Human zin
399	39	45.3	890	6	ABU15116	Abu15116	Protein e	472	38	44.2	313	4	AAB90653	Aab90653	Human sec
400	39	45.3	959	8	ABO84640	AbO84640	Human can	473	38	44.2	326	4	AAB90652	Aab90652	Human sec
401	39	45.3	960	8	ABO84643	AbO84643	Human can	474	38	44.2	353	8	ADS44665	Ads44665	Bacterial
402	39	45.3	971	7	ADD46684	Add46684	Rat Prote	475	38	44.2	385	3	AAB19189	Aab19189	Lipid met
403	39	45.3	986	8	ABO84645	AbO84645	Human can	476	38	44.2	385	9	ADV70586	Adv70586	Brevibact
404	39	45.3	1007	7	ADD01193	Add01193	Human nuc	477	38	44.2	385	9	ADW15128	Adw15128	B. lactof
405	39	45.3	1013	8	ABO84644	AbO84644	Human can	478	38	44.2	395	8	ADS21844	Ads21844	Bacterial
406	39	45.3	1017	8	ABO84642	AbO84642	Human can	479	38	44.2	398	8	ADS41600	Ads41600	Bacterial
407	39	45.3	1018	8	ABO84641	AbO84641	Human can	480	38	44.2	402	4	ABB50185	Abb50185	Human tra
408	39	45.3	1113	6	ABB82999	Abb82999	Human KCN	481	38	44.2	402	8	ADJ64309	Adj64309	Cartilage
409	39	45.3	1146	3	AY92225	Ay92225	Human pat	482	38	44.2	402	9	AED26270	Aed26270	Novel hum
410	39	45.3	1154	6	ABO07240	AbO07240	Human p53	483	38	44.2	403	2	AAY21963	Aay21963	Mycobacte
411	39	45.3	1154	6	ABB83000	Abb83000	Human KCN	484	38	44.2	403	2	AAW72943	Aaw72943	Mycobacte
412	39	45.3	1154	7	ADD46686	Add46686	Human Pro	485	38	44.2	404	2	AAW72942	Aaw72942	Mycobacte
413	39	45.3	1154	8	ADJ75670	Adj75670	Marker ge	486	38	44.2	404	2	AAW72942	Aaw72942	Mycobacte
414	39	45.3	1171	8	ADJ76368	Adj76368	Marker ge	487	38	44.2	410	8	ADJ64307	Adj64307	Cartilage
415	39	45.3	1171	8	ABO84639	AbO84639	Mouse can	488	38	44.2	419	4	ABB66861	Abb66861	Drosophil
416	39	45.3	1172	8	ADO59149	Ado59149	Chicken S	489	38	44.2	424	4	ABB611372	Abb611372	Drosophil
417	39	45.3	1174	7	ABM79021	Abm79021	BK channe	490	38	44.2	483	5	ABB76634	Abb76634	Ternamyl-
418	39	45.3	1178	8	ABO84648	AbO84648	Human can	491	38	44.2	483	5	ABB76611	Abb76611	Ternamyl-
419	39	45.3	1196	2	AAV32017	Aay32017	Mouse cat	492	38	44.2	483	5	ABB76632	Abb76632	Ternamyl-
420	39	45.3	1196	2	ADJ38331	Adi38331	Mouse cat	493	38	44.2	483	5	ABB76622	Abb76622	Ternamyl-
421	39	45.3	1196	10	AE668557	Aee68557	Mouse pot	494	38	44.2	483	5	ABB76624	Abb76624	Ternamyl-
422	39	45.3	1203	2	AAV43261	Aay43261	Human pat	495	38	44.2	483	5	ABB76616	Abb76616	Ternamyl-
423	39	45.3	1203	2	AAV28444	Aay28444	Human ptc	496	38	44.2	483	5	ABB76612	Abb76612	Ternamyl-
424	39	45.3	1203	3	AAV92703	Aay92703	Human pat	497	38	44.2	483	5	ABB76613	Abb76613	Ternamyl-
425	39	45.3	1203	5	AAE19829	Aae19829	Human pat	498	38	44.2	483	5	ABB76615	Abb76615	Ternamyl-
426	39	45.3	1203	6	ABG74104	Abg74104	Human pat	499	38	44.2	483	5	ABB76631	Abb76631	Ternamyl-
427	39	45.3	1207	8	ABO84646	AbO84646	Human can	500	38	44.2	483	5	ABB76614	Abb76614	Ternamyl-
428	39	45.3	1236	8	ABO84647	AbO84647	Human can	501	38	44.2	483	5	ABB76623	Abb76623	Ternamyl-
429	39	45.3	1323	7	ADD25207	Add25207	Fertility	502	38	44.2	483	5	ABB76633	Abb76633	Ternamyl-
430	39	45.3	1323	8	ADN61222	Adn61222	Radish nu	503	38	44.2	483	5	ABB76618	Abb76618	Ternamyl-
431	39	45.3	1331	4	AAW39048	Aam39048	Human pol	504	38	44.2	483	5	ABB76625	Abb76625	Ternamyl-
432	39	45.3	1516	8	ADQ97109	Adq97109	Mouse can	505	38	44.2	483	5	ABB76629	Abb76629	Ternamyl-
433	39	45.3	1927	7	ADDA48256	Adm48256	Human pro	506	38	44.2	483	5	ABB76617	Abb76617	Ternamyl-
434	39	45.3	1933	8	ABM82906	Abm82906	Human dia	507	38	44.2	483	5	ABB76630	Abb76630	Ternamyl-
435	38.5	44.8	366	8	ADX70942	Adx70942	Plant ful	508	38	44.2	488	4	AAB31471	Aab31471	Amino aci
436	38.5	44.8	480	8	ADL99494	Adl99494	Human leu	509	38	44.2	488	4	AAB31469	Aab31469	Amino aci
437	38.5	44.8	532	8	ADL99491	Adl99491	Human leu	510	38	44.2	488	4	AAB31472	Aab31472	Amino aci
438	38.5	44.8	563	5	ABBA48302	Abb48302	Listeria	511	38	44.2	488	4	AAB31470	Aab31470	Amino aci
439	38.5	44.8	601	10	AEF72836	Aef72836	Human leu	512	38	44.2	502	8	ADY11010	Ady11010	Plant ful
440	38.5	44.8	610	8	ADT66633	Adt66633	Rat leuko	513	38	44.2	519	6	ABP98897	Abp98897	Human mol
441	38.5	44.8	611	3	AAB08840	Aab08840	Amino aci	514	38	44.2	607	2	AAW09422	Aaw09422	Banana po
442	38.5	44.8	611	7	ADBE25724	Adbe25724	Human pro	515	38	44.2	645	8	ADJ93708	Adj93708	ACH2 pro
443	38.5	44.8	611	9	ABM81460	Abm81460	Tumour-as	516	38	44.2	647	4	ABB60060	Abb60060	Drosophil
444	38.5	44.8	611	9	ADW07254	Adw07254	Human hep	517	38	44.2	649	4	ABB66538	Abb66538	Drosophil
445	38.5	44.8	611	9	ADZ04287	Adz04287	Human leu	518	38	44.2	658	4	ABG27820	Abg27820	Novel hum
446	38.5	44.8	625	3	AAB58111	Aab58111	Lung can	519	38	44.2	773	9	AEC08348	Aec08348	A. evansi
447	38	44.2	14	AAE12159	Aae12159	Human hg-	520	38	44.2	798	4	AAU03530	Aau03530	Human pro	
448	38	44.2	60	4	AAE12157	Aae12157	Human hg-	521	38	44.2	798	4	AAB85787	Ab85787	Human kin
449	38	44.2	128	6	ADA83800	Ada83800	Human FAB	522	38	44.2	804	8	ADX66686	Adx66686	Plant ful
450	38	44.2	140	3	AG01491	Ag01491	Human sec	523	38	44.2	891	4	ABB64847	Abb64847	Drosophil
451	38	44.2	152	4	AAU32657	Aau32657	Novel hum	524	38	44.2	904	6	ADB17511	Adb17511	Rice post
452	38	44.2	182	6	ABU44226	Abu44226	Protein e	525	38	44.2	904	9	AEC75735	Aec75735	Rice japo
453	38	44.2	185	4	ABH69211	Abh69211	Drosophil	526	38	44.2	915	7	ABO61732	Abo61732	Klebsiell
454	38	44.2	186	4	ABU53118	Abu53118	Intracell	527	38	44.2	916	6	ADB17498	Adb17498	Wheat pos
455	38	44.2	192	8	ADX72427	Adx72427	Plant ful	528	38	44.2	916	9	AEC75722	Aec75722	Wheat Arg
456	38	44.2	195	8	AEC40516	Aec40516	Deltan50	529	38	44.2	932	6	ABU16693	Abu16693	Protein e
457	38	44.2	207	9	AEC40515	Aec40515	Deltan40	530	38	44.2	964	2	AAW32619	Aaw32619	Cyclic-is
458	38	44.2	210	2	AAW23087	Aaw23087	Polyangiu	531	38	44.2	1005	4	ABG11447	Abg11447	Novel hum
459	38	44.2	222	9	AEC40514	Aec40514	Deltan20	532	38	44.2	1139	8	ADR73937	Adr73937	Human man
460	38	44.2	248	8	ADX68808	Adx68808	Plant ful	533	38	44.2	1139	9	ABE71164	Aeb71164	Alpha-man
461	38	44.2	248	9	AEC40497	Aec40497	BFDV isol	534	38	44.2	1180	4	AAU32658	Aau32658	Novel hum

535	38	44.2	1939	4	ABG20505	Abg20505 Novel hum	608	37	43.0	269	9	AEA58368	Aea58368 Streptoco
536	38	44.2	2017	4	AAU32654	Aau32654 Novel hum	609	37	43.0	273	7	ADC18688	Adc18688 Mouse gp
537	38	44.2	2202	4	AAU32655	Aau32655 Novel hum	610	37	43.0	275	7	ADC18689	Adc18689 Mouse gp
538	38	44.2	2858	4	ABB71150	Abb71150 Drosophil	611	37	43.0	276	7	ADC18687	Adc18687 Mouse gp
539	38	44.2	3060	4	ABB58064	Abb58064 Drosophil	612	37	43.0	280	7	ABM86043	Abm86043 Rice abio
540	37.5	43.6	41	2	AAW33674	Aaw33674 FIV PPR C	613	37	43.0	283	8	ADK47844	Adk47844 Streptoco
541	37.5	43.6	288	9	ADX40464	Adx40464 HIV Vif p	614	37	43.0	297	4	ABG15100	Abg15100 Novel hum
542	37.5	43.6	347	8	ABM84252	Abm84252 Human dia	615	37	43.0	310	10	AEF54140	Aef54140 His tagge
543	37.5	43.6	420	8	ABM84251	Abm84251 Human dia	616	37	43.0	315	7	ADD12568	Add12568 Human ENZ
544	37.5	43.6	446	8	ABM84250	Abm84250 Human dia	617	37	43.0	316	7	ABO81579	AbO81579 Pseudomon
545	37.5	43.6	449	7	ADC78253	Adc78253 Human sec	618	37	43.0	319	9	ABE42107	Aeb42107 L. pneumo
546	37.5	43.6	449	9	AED11499	Aed11499 Human gen	619	37	43.0	323	9	ABE38947	Aeb38947 L. pneumo
547	37.5	43.6	546	5	ABB57161	Abb57161 Mouse iSc	620	37	43.0	332	8	ADY12827	Ady12827 plant ful
548	37.5	43.6	550	8	ADM33966	Adm33966 Human DGC	621	37	43.0	374	8	ADS27118	Ads27118 Bacterial
549	37.5	43.6	550	9	ADX06171	Adx06171 Cyclin-de	622	37	43.0	374	8	ADS26366	Ads26366 Bacterial
550	37.5	43.6	576	8	ABM84248	Abm84248 Human dia	623	37	43.0	374	8	ADN21639	Adn21639 Bacterial
551	37.5	43.6	616	8	ABM84249	Abm84249 Human dia	624	37	43.0	375	8	ADN21639	Adn21639 Bacterial
552	37	43.0	8	3	AAV67360	Aav67360 Melanoma	625	37	43.0	375	8	ADN24395	Adn24395 Bacterial
553	37	43.0	9	3	AAO3467	Aao3467 Human tyr	626	37	43.0	379	8	ADS43401	Ads43401 Bacterial
554	37	43.0	9	3	AAO3474	Aao3474 Human tyr	627	37	43.0	388	8	ADN21799	Adn21799 Bacterial
555	37	43.0	9	4	AAU26665	Aau26665 Human Leu	628	37	43.0	388	8	ADN24559	Adn24559 Bacterial
556	37	43.0	9	4	AAU26998	Aau26998 Human Leu	629	37	43.0	389	6	ABU21323	Abu21323 Protein e
557	37	43.0	9	4	AAU26997	Aau26997 Human Leu	630	37	43.0	395	6	ABU21149	Abu21149 Protein e
558	37	43.0	9	4	AAU26664	Aau26664 Human Leu	631	37	43.0	408	4	ABO96440	AbO96440 Putative
559	37	43.0	9	5	AAE31324	Aae31324 Human tyr	632	37	43.0	420	7	ABO76672	AbO76672 Pseudomon
560	37	43.0	9	5	AAE31384	Aae31384 Human Tyr	633	37	43.0	431	7	ADC99069	Adc99069 Human KPP
561	37	43.0	9	5	AAE31325	Aae31325 Human tyr	634	37	43.0	443	9	ADY39715	Ady39715 MADS box
562	37	43.0	9	7	ADN14217	Adn14217 Human mel	635	37	43.0	445	8	ADN25882	Adn25882 Bacterial
563	37	43.0	9	7	ADW31578	Adw31578 HLA bindi	636	37	43.0	447	9	AEA81265	Aea81265 Human gal
564	37	43.0	10	4	AAU26718	Aau26718 Human Leu	637	37	43.0	448	3	AAV71285	Aav71285 Streptoco
565	37	43.0	10	4	AAU26707	Aau26707 Human Leu	638	37	43.0	448	6	ABU02732	Abu02732 S. pneumo
566	37	43.0	10	4	AAU27051	Aau27051 Human Leu	639	37	43.0	448	8	ADM92272	Adm92272 S. pneumon
567	37	43.0	10	4	AAU27040	Aau27040 Human Leu	640	37	43.0	448	8	ADT50163	Adt50163 S. pneumon
568	37	43.0	23	2	AAV36451	Aav36451 Fragment	641	37	43.0	458	9	AEA81266	Aea81266 Human aci
569	37	43.0	23	6	ADA11986	Ada11986 Human nov	642	37	43.0	467	2	AAV35800	Aav35800 Amino aci
570	37	43.0	27	10	AEF54139	Aef54139 Histidine	643	37	43.0	467	2	AAV35800	Aav35800 Amino aci
571	37	43.0	30	5	AAU84938	Aau84938 Human Trp	644	37	43.0	468	8	ABM83258	Abm83258 Human dia
572	37	43.0	30	5	AAU84873	Aau84873 Human Trp	645	37	43.0	488	8	ABM83259	Abm83259 Human dia
573	37	43.0	32	2	AAU97388	Aau97388 CC49 VH-8	646	37	43.0	491	3	RAAB36459	Raa36459 Lemon acy
574	37	43.0	41	4	ABBI17688	Abbi17688 Human ner	647	37	43.0	499	7	ADG70544	Adg70544 Aspergill
575	37	43.0	51	5	ABP63707	Abp63707 Human ORF	648	37	43.0	500	8	ADS42425	Ads42425 Bacterial
576	37	43.0	52	4	AAU62413	Aau62413 Propionib	649	37	43.0	502	6	ABU63509	Abu63509 Ornithine
577	37	43.0	52	6	ABM58932	Abm58932 Propionib	650	37	43.0	517	2	AAV31981	Aav31981 Mouse tyr
578	37	43.0	53	4	AAU06399	Aau06399 Human foe	651	37	43.0	517	2	AAV42635	Aav42635 Murine tyr
579	37	43.0	62	2	AAW20936	Aaw20936 H. pylori	652	37	43.0	517	5	AAU76663	Aau76663 Mouse tyr
580	37	43.0	67	4	AAU46194	Aau46194 Propionib	653	37	43.0	517	6	ABR82196	AbR82196 Tyrosinas
581	37	43.0	67	6	ABM42713	Abm42713 Propionib	654	37	43.0	519	2	AAW30826	Aaw30826 The novel
582	37	43.0	80	4	AAU63600	Aau63600 Propionib	655	37	43.0	519	2	AAV31982	Aav31982 Human tyr
583	37	43.0	80	6	ABM60119	Abm60119 Propionib	656	37	43.0	519	4	AAV42636	Aav42636 Human tyr
584	37	43.0	85	2	AAV36449	Aav36449 Fragment	657	37	43.0	519	4	AAAB86041	AaB86041 Human Trp
585	37	43.0	85	6	ADA11984	Ada11984 Human nov	658	37	43.0	519	5	AAU84807	Aau84807 Human Trp
586	37	43.0	86	4	ADS08142	Ads08142 Staphyloc	659	37	43.0	519	5	ADI28007	Adi28007 Human tyr
587	37	43.0	96	4	AAO12637	Aao12637 Human pol	660	37	43.0	519	6	ABR58675	AbR58675 Human can
588	37	43.0	107	5	ABP03356	Abp03356 Human ORF	661	37	43.0	519	8	ADM12479	Adm12479 Human dop
589	37	43.0	110	2	AAV21366	Aav21366 Human HUP	662	37	43.0	519	8	ADO38702	AdO38702 Tyrosinas
590	37	43.0	111	6	ABR64186	AbR64186 Angiogene	663	37	43.0	519	8	ADQ19864	Adq19864 Human sof
591	37	43.0	116	8	ADX92978	Adx92978 Plant ful	664	37	43.0	519	9	ADY85097	AdY85097 Tumor ant
592	37	43.0	127	2	AAW20574	Aaw20574 H. pylori	665	37	43.0	519	10	AEF01187	Aef01187 Tyrosinas
593	37	43.0	127	7	ADE57168	Ade57168 Human Pro	666	37	43.0	537	4	AAAB86040	AaB86040 Human Trp
594	37	43.0	157	7	ABO74821	AbO74821 Pseudomon	667	37	43.0	537	5	ADI28006	Adi28006 Human tyr
595	37	43.0	181	8	ADN18913	Adn18913 Bacterial	668	37	43.0	546	8	ADN17427	Adn17427 Bacterial
596	37	43.0	181	8	ADN18897	Adn18897 Bacterial	669	37	43.0	551	4	ABAB48829	Aab48829 Thermotog
597	37	43.0	182	3	ABN29563	Abn29563 Trypanoso	670	37	43.0	552	5	ABO7927	AbO7927 T. mariti
598	37	43.0	198	4	ABG30053	Abg30053 Novel hum	671	37	43.0	552	6	ABB82547	AbB82547 Human tyr
599	37	43.0	223	8	ADV82139	Adv82139 Streptoco	672	37	43.0	552	8	ADRS1311	AdRS1311 Anti-biof
600	37	43.0	225	5	ABP25844	Abp25844 Streptoco	673	37	43.0	554	8	ADS23691	Ads23691 Bacterial
601	37	43.0	225	8	ADV88748	Adv88748 Streptoco	674	37	43.0	555	2	AAW34563	Aaw34563 Thermotog
602	37	43.0	225	8	ADV80001	Adv80001 Streptoco	675	37	43.0	555	2	AAW49867	Aaw49867 Thermotog
603	37	43.0	234	2	AAE60794	Aae60794 Kawasaki	676	37	43.0	555	7	ADN73254	Adn73254 Protein r
604	37	43.0	248	4	AAAB24180	AaB24180 Pseudomon	677	37	43.0	555	7	ADC26915	AdC26915 Thermotog
605	37	43.0	248	5	AAU84805	Aau84805 Human TRP	678	37	43.0	555	7	ADE93811	AdE93811 T. mariti
606	37	43.0	249	2	AAAR14258	Aar14258 gp75 pept	679	37	43.0	555	9	ADW95037	AdW95037 Torenia h
607	37	43.0	269	8	ADR94498	Adr94498 Novel S.	680	37	43.0	555	9	ADX05239	AdX05239 Torenia h

681	37	43.0	576	3	AAV75498	Aay75498 Neisseria	754	36	41.9	9	9	ADU70736	AdU70736 Human hep
682	37	43.0	576	6	ABP80808	N. gonorr	755	36	41.9	11	10	AEES59474	Aee59474 Natriuret
683	37	43.0	618	2	AAW34991	Aaw34991 Thermotog	756	36	41.9	13	10	AEES59467	Aee59467 Natriuret
684	37	43.0	684	4	ABG24375	Novel hum	757	36	41.9	15	9	ADU71289	AdU71289 Human hep
685	37	43.0	751	4	AAE09958	Methylomo	758	36	41.9	17	3	ABAB06401	Aab06401 Randomise
686	37	43.0	751	5	ABG61579	High grow	759	36	41.9	48	8	ADJ12358	AdJ12358 Peptide f
687	37	43.0	751	6	ABP77444	N. gonorr	760	36	41.9	61	4	AAU48383	Aau48383 Propionib
688	37	43.0	773	3	ABAB08200	Amino aci	761	36	41.9	61	5	ABP01082	Abp01082 Human ORF
689	37	43.0	773	9	ADW23592	Human dia	762	36	41.9	61	6	ABM44902	Abm44902 Propionib
690	37	43.0	773	10	AEF14405	Human cho	763	36	41.9	63	6	ABM65503	Abm65503 Propionib
691	37	43.0	796	6	ABU17006	Protein e	764	36	41.9	65	4	AAU81868	Aau81868 Human hae
692	37	43.0	797	5	AAE19149	Human kin	765	36	41.9	65	4	AAU81868	Aau81868 Human hae
693	37	43.0	804	3	ABAB08199	Amino aci	766	36	41.9	97	4	AAU54279	Aau54279 Propionib
694	37	43.0	804	8	ADJ96665	Human lip	767	36	41.9	97	6	ABM50798	Abm50798 Propionib
695	37	43.0	804	10	AEF14404	Human cho	768	36	41.9	100	4	AAU67296	Aau67296 Propionib
696	37	43.0	869	5	ABP65968	Bifidobac	769	36	41.9	100	6	ABM63815	Abm63815 Propionib
697	37	43.0	952	4	ABB69760	Drosophil	770	36	41.9	100	7	ABO63532	Abo63532 Klebsiell
698	37	43.0	1023	4	ABG24378	Novel hum	771	36	41.9	113	4	AAU51623	Aau51623 Propionib
699	37	43.0	1032	6	ABU15004	Protein e	772	36	41.9	113	6	ABM48142	Abm48142 Propionib
700	37	43.0	1043	4	ABBS8772	Drosophil	773	36	41.9	113	6	ABJ25288	Abj25288 Mouse bac
701	37	43.0	1156	8	ADS42937	Bacterial	774	36	41.9	115	8	ADP30206	Adp30206 Human sec
702	37	43.0	1201	8	ADS30065	Bacterial	775	36	41.9	116	4	AAU63310	Aau63310 Propionib
703	37	43.0	1368	5	ABP73707	Candida a	776	36	41.9	116	6	ABM59829	Abm59829 Propionib
704	37	43.0	1368	5	ADJ28002	Human TRP	777	36	41.9	121	3	AAU20409	Aau20409 Arabidops
705	37	43.0	1384	5	ADJ28004	Human TRP	778	36	41.9	122	5	ABP11305	Abp11305 Human ORF
706	37	43.0	1392	5	ADJ28000	Human TRP	779	36	41.9	131	8	ADK71682	Adk71682 Pseudomon
707	37	43.0	1605	4	ABP70375	Drosophil	780	36	41.9	135	7	ABO78076	Abo78076 Pseudomon
708	37	43.0	1674	9	AEBS2335	Human pro	781	36	41.9	158	5	ABB97984	Abb97984 Human cau
709	37	43.0	1674	9	AEBS2340	Human pro	782	36	41.9	168	8	ADK71841	Adk71841 Human kin
710	37	43.0	1793	4	ABG25611	Novel hum	783	36	41.9	178	6	ABU34183	Abu34183 Protein e
711	37	43.0	2071	6	ABU04321	Human exp	788	36	41.9	188	8	ADT55527	Adt55527 Plant pol
712	37	43.0	2338	4	ABU04312	Human exp	789	36	41.9	203	4	AAU20408	Aau20408 Arabidops
713	37	43.0	2491	5	AAE24315	Human cat	785	36	41.9	184	4	ABU53120	Abu53120 Intracell
714	37	43.0	2491	6	ABU04319	Human exp	786	36	41.9	187	7	ABO73380	Abo73380 Pseudomon
715	37	43.0	2491	6	ABU04321	Human exp	787	36	41.9	198	3	AAU20408	Aau20408 Arabidops
716	37	43.0	2491	6	ABU04315	Human exp	788	36	41.9	198	8	ADT55527	Adt55527 Plant pol
717	37	43.0	2491	6	ABU04314	Human exp	790	36	41.9	203	4	AAU20408	Aau20408 Arabidops
718	37	43.0	2491	6	ABU04317	Human exp	791	36	41.9	203	8	ADT55527	Adt55527 Plant pol
719	37	43.0	2491	6	ABU04316	Human exp	792	36	41.9	211	8	ADT55527	Adt55527 Plant pol
720	37	43.0	2491	6	ABU04313	Human exp	793	36	41.9	219	4	AAU62002	Aau62002 Propionib
721	37	43.0	2491	6	ABU04320	Human exp	794	36	41.9	219	6	ABM58521	Abm58521 Propionib
722	37	43.0	2491	7	ADJ68259	Human exp	795	36	41.9	225	8	ADY06493	Ady06493 Plant ful
723	37	43.0	2491	8	ADJ68259	Human exp	796	36	41.9	225	8	ADY06493	Ady06493 Plant ful
724	37	43.0	2491	8	ADJ68259	Human exp	797	36	41.9	226	4	AAU23742	Aau23742 Human EST
725	37	43.0	2491	8	ADJ68259	Human exp	798	36	41.9	226	4	ABG02211	Abg02211 Novel hum
726	37	43.0	2491	8	ADJ68259	Human exp	799	36	41.9	226	4	ABG02211	Abg02211 Novel hum
727	37	43.0	2491	9	ADJ68259	Human exp	800	36	41.9	234	3	AAU23340	Aau23340 Protein e
728	37	43.0	2491	9	ADJ68259	Human exp	801	36	41.9	234	6	ABU23340	Abu23340 Protein e
729	37	43.0	2491	9	ADJ68259	Human exp	802	36	41.9	234	8	ADT55527	Adt55527 Plant pol
730	37	43.0	2491	10	ADJ68259	Human exp	803	36	41.9	234	8	ADT55527	Adt55527 Plant pol
731	37	43.0	2491	10	ADJ68259	Human exp	804	36	41.9	234	8	ADT55527	Adt55527 Plant pol
732	37	43.0	2491	10	ADJ68259	Human exp	805	36	41.9	234	8	ADT55527	Adt55527 Plant pol
733	37	43.0	2491	10	ADJ68259	Human exp	806	36	41.9	234	8	ADT55527	Adt55527 Plant pol
734	37	43.0	2491	10	ADJ68259	Human exp	807	36	41.9	234	8	ADT55527	Adt55527 Plant pol
735	36.5	42.4	101	4	AAU03669	Human pol	808	36	41.9	234	8	ADT55527	Adt55527 Plant pol
736	36.5	42.4	157	6	ABU09576	VEE virus	809	36	41.9	234	8	ADT55527	Adt55527 Plant pol
737	36.5	42.4	165	6	ABU09577	Mutant VE	810	36	41.9	234	8	ADT55527	Adt55527 Plant pol
738	36.5	42.4	165	6	ABU09578	Mutant VE	811	36	41.9	234	8	ADT55527	Adt55527 Plant pol
739	36.5	42.4	169	6	ABU09579	TatCCD (C	812	36	41.9	234	8	ADT55527	Adt55527 Plant pol
740	36.5	42.4	191	10	AEF63481	Modified	813	36	41.9	234	8	ADT55527	Adt55527 Plant pol
741	36.5	42.4	192	8	ADP20069	Human imm	814	36	41.9	234	8	ADT55527	Adt55527 Plant pol
742	36.5	42.4	192	9	ADX40511	HIV Vif p	815	36	41.9	234	8	ADT55527	Adt55527 Plant pol
743	36.5	42.4	192	9	ADX40520	HIV Vif p	816	36	41.9	234	8	ADT55527	Adt55527 Plant pol
744	36.5	42.4	192	9	ADX40518	HIV Vif p	817	36	41.9	234	8	ADT55527	Adt55527 Plant pol
745	36.5	42.4	192	9	ADX40512	HIV Vif p	818	36	41.9	234	8	ADT55527	Adt55527 Plant pol
746	36.5	42.4	192	9	ADX40510	HIV Vif p	819	36	41.9	234	8	ADT55527	Adt55527 Plant pol
747	36.5	42.4	268	5	ADP25956	Streptoco	820	36	41.9	234	8	ADT55527	Adt55527 Plant pol
748	36.5	42.4	268	8	ADR83944	S. pyogen	821	36	41.9	234	8	ADT55527	Adt55527 Plant pol
749	36.5	42.4	288	9	ADX40444	HIV Vif p	822	36	41.9	234	8	ADT55527	Adt55527 Plant pol
750	36.5	42.4	342	5	AAE18305	venezuela	823	36	41.9	234	8	ADT55527	Adt55527 Plant pol
751	36.5	42.4	381	7	ABM88307	Rice abio	824	36	41.9	234	8	ADT55527	Adt55527 Plant pol
752	36.5	42.4	696	7	ABO79890	Pseudomon	825	36	41.9	234	8	ADT55527	Adt55527 Plant pol
753	36.5	42.4	1439	10	ABF63466	pol-linke	826	36	41.9	234	8	ADT55527	Adt55527 Plant pol

827	36	41.9	383	9	AEb37429	Aeb37429	L. pneumo	900	36	41.9	485	3	AAB29363	Aab29363	Bacillus
828	36	41.9	395	6	ABU17339	Abu17339	Protein e	901	36	41.9	485	3	AAB29387	Aab29387	Bacillus
829	36	41.9	391	6	ADA33574	Ada33574	Acinetoba	902	36	41.9	485	3	AAB29389	Aab29389	Bacillus
830	36	41.9	396	8	ADY13837	Ady13837	Bacterial	903	36	41.9	485	3	AAB29385	Aab29385	Bacillus
831	36	41.9	423	8	ADN22904	Adn22904	Propionib	904	36	41.9	485	3	AAB29365	Aab29365	Bacillus
832	36	41.9	426	6	ABM65116	Abm65116	Rice abio	905	36	41.9	485	3	AAB29386	Aab29386	Bacillus
833	36	41.9	429	7	ABM88474	Abm88474	Human pro	906	36	41.9	485	3	AAB29388	Aab29388	Bacillus
834	36	41.9	430	6	ADA54609	Ada54609	Human pro	907	36	41.9	485	3	AAB29394	Aab29394	Bacillus
835	36	41.9	441	8	ABY00045	AbY00045	Porcine e	908	36	41.9	485	3	AAB29366	Aab29366	Bacillus
836	36	41.9	443	4	ABG14210	Abg14210	Novel hum	909	36	41.9	488	9	AED19768	Aed19768	Baculovir
837	36	41.9	447	6	ADA55102	Ada55102	Human pro	910	36	41.9	491	9	AED19774	Aed19774	Baculovir
838	36	41.9	449	6	ABU18372	Abu18372	Protein e	911	36	41.9	502	3	AAG06850	Aag06850	Arabidops
839	36	41.9	450	3	AAG41505	Aag41505	Arabidops	912	36	41.9	511	2	AAW78475	Aaw78475	Autograph
840	36	41.9	450	5	ABB93654	Abb93654	Herbicida	913	36	41.9	512	8	ADG16321	Adg16321	Antibody
841	36	41.9	452	3	AAG06851	Aag06851	Arabidops	914	36	41.9	512	8	AED19766	Aed19766	Baculovir
842	36	41.9	453	8	ADN20522	Adn20522	Bacterial	915	36	41.9	517	5	AAE25051	Aae25051	GP64-6His
843	36	41.9	455	3	AAG41504	Aag41504	Arabidops	916	36	41.9	518	8	ADG16323	Adg16323	Antibody
844	36	41.9	458	3	AAG41503	Aag41503	Arabidops	917	36	41.9	518	8	ADT04559	Adt04559	Peptide t
845	36	41.9	463	3	AAG06588	Aag06588	Arabidops	918	36	41.9	528	8	ABM83086	Abm83086	Human dia
846	36	41.9	463	3	AAG48388	Aag48388	Arabidops	919	36	41.9	530	2	AAW78476	Aaw78476	Baculovir
847	36	41.9	463	5	ABB93655	Abb93655	Herbicida	920	36	41.9	532	2	AAI39296	Aai39296	Alkaline
848	36	41.9	463	6	ABU36773	Abu36773	Protein e	921	36	41.9	532	8	ABM83085	Abm83085	Human dia
849	36	41.9	470	3	AAG06587	Aag06587	Arabidops	922	36	41.9	538	9	ADV14310	Adv14310	Corn Na+/
850	36	41.9	470	3	AAG48387	Aag48387	Arabidops	923	36	41.9	538	9	ADW95062	Adw95062	Zea maye
851	36	41.9	481	2	AAW55029	Aaw55029	G-protein	924	36	41.9	538	9	ADX05264	Adx05264	Zea maye
852	36	41.9	481	2	AAI25969	Aai25969	Human bra	925	36	41.9	540	4	ABB64912	Abb64912	Drosophi
853	36	41.9	481	2	AAI25969	Aai25969	Human bra	926	36	41.9	542	2	AAW55030	Aaw55030	G-protein
854	36	41.9	481	3	AAI25969	Aai25969	Human bra	927	36	41.9	545	8	ABM83087	Abm83087	Human dia
855	36	41.9	481	3	AAI25969	Aai25969	Human bra	928	36	41.9	558	8	ADU07866	Adu07866	Amino aci
856	36	41.9	481	3	AAI25969	Aai25969	Human bra	929	36	41.9	576	8	ABM83084	Abm83084	Human dia
857	36	41.9	481	5	ABG95173	Abg95173	Human GPC	930	36	41.9	587	4	ABM67655	Abm67655	Amino aci
858	36	41.9	481	5	ABG95160	Abg95160	Human GPC	931	36	41.9	587	8	ABM83083	Abm83083	Human dia
859	36	41.9	481	6	ABP82006	Abp82006	Human end	932	36	41.9	587	8	ADU07900	Adu07900	Amino aci
860	36	41.9	481	6	ABP82006	Abp82006	Human end	933	36	41.9	587	8	ADU07904	Adu07904	Amino aci
861	36	41.9	481	7	ADB67657	Adb67657	Human end	934	36	41.9	587	8	ADW76538	Adw76538	Alkaliph
862	36	41.9	481	7	ADC22601	Adc22601	Human G p	935	36	41.9	590	8	ADU07894	Adu07894	Amino aci
863	36	41.9	481	7	ADC22727	Adc22727	Human G p	936	36	41.9	590	8	ADU07912	Adu07912	Amino aci
864	36	41.9	481	7	ADH14200	Adh14200	Mutated h	937	36	41.9	591	8	ABM83082	Abm83082	Human dia
865	36	41.9	481	7	ADH14074	Adh14074	Human ETB	938	36	41.9	601	8	ABM83081	Abm83081	Human dia
866	36	41.9	481	8	ADO29058	Ado29058	Human nov	939	36	41.9	616	8	ABM83089	Abm83089	Human dia
867	36	41.9	481	8	ABM80229	Abm80229	Tumour-ag	940	36	41.9	671	6	ABR52866	AbR52866	Protein s
868	36	41.9	481	9	AED34533	Aed34533	Human ETB	941	36	41.9	671	7	ADK62338	Adk62338	Disease t
869	36	41.9	482	8	ADO29059	Ado29059	Mouse nov	942	36	41.9	671	8	ADN19334	Adn19334	Bacterial
870	36	41.9	483	3	AAB29370	Aab29370	Bacillus	943	36	41.9	672	8	ABM83080	Abm83080	Human dia
871	36	41.9	483	3	AAB29372	Aab29372	Bacillus	944	36	41.9	683	8	ADJ75550	Adj75550	Marker ge
872	36	41.9	483	3	AAB29380	Aab29380	Bacillus	945	36	41.9	685	8	ADR39739	Adr39739	Human kin
873	36	41.9	483	3	AAB29371	Aab29371	Bacillus	946	36	41.9	698	4	ABG08662	Abg08662	Novel hum
874	36	41.9	483	3	AAB29376	Aab29376	Bacillus	947	36	41.9	711	8	ABM83079	Abm83079	Human dia
875	36	41.9	483	3	AAB29325	Aab29325	Bacillus	948	36	41.9	715	6	ABU03523	Abu03523	Angiogene
876	36	41.9	483	3	AAB29384	Aab29384	Bacillus	949	36	41.9	715	8	ADN03820	Adn03820	Antipeori
877	36	41.9	483	3	AAB29328	Aab29328	Bacillus	950	36	41.9	715	8	ADO19246	Ado19246	Human PRO
878	36	41.9	483	3	AAB29377	Aab29377	Bacillus	951	36	41.9	715	8	ABM83078	Abm83078	Human dia
879	36	41.9	483	3	AAB29378	Aab29378	Bacillus	952	36	41.9	715	8	ABM83077	Abm83077	Human dia
880	36	41.9	483	3	AAB29382	Aab29382	Bacillus	953	36	41.9	716	8	ABM83076	Abm83076	Human dia
881	36	41.9	483	3	AAB29374	Aab29374	Bacillus	954	36	41.9	719	7	ADF70494	Adf70494	Orphan re
882	36	41.9	483	3	AAB29379	Aab29379	Bacillus	955	36	41.9	735	7	ADE76568	Ade76568	Adeno-ass
883	36	41.9	483	3	AAB29369	Aab29369	Bacillus	956	36	41.9	735	7	ADE76569	Ade76569	Adeno-ass
884	36	41.9	483	3	AAB29373	Aab29373	Bacillus	957	36	41.9	735	7	ADE76567	Ade76567	Adeno-ass
885	36	41.9	483	3	AAB29381	Aab29381	Bacillus	958	36	41.9	735	9	ADZ27084	Adz27084	Adeno-ass
886	36	41.9	483	3	AAB29375	Aab29375	Bacillus	959	36	41.9	740	8	ABM83075	Abm83075	Human dia
887	36	41.9	483	3	AAB29329	Aab29329	Bacillus	960	36	41.9	740	8	ABM83074	Abm83074	Human dia
888	36	41.9	483	3	AAB29330	Aab29330	Bacillus	961	36	41.9	759	8	ABM83088	Abm83088	Human dia
889	36	41.9	483	3	AAB29383	Aab29383	Bacillus	962	36	41.9	784	8	ADS30331	Ads30331	Bacterial
890	36	41.9	484	6	ABU34698	Abu34698	Protein e	963	36	41.9	785	9	ABM90808	Abm90808	M. xanthu
891	36	41.9	485	3	AAB29364	Aab29364	Bacillus	964	36	41.9	840	9	ADM00948	Adm00948	Amino aci
892	36	41.9	485	3	AAB29392	Aab29392	Bacillus	965	36	41.9	845	8	ADY23636	Ady23636	Plant ful
893	36	41.9	485	3	AAB29327	Aab29327	Bacillus	966	36	41.9	907	8	ADP29998	Adp29998	Human sec
894	36	41.9	485	3	AAB29390	Aab29390	Bacillus	967	36	41.9	912	4	AAU52873	Aau52873	Propionib
895	36	41.9	485	3	AAB29393	Aab29393	Bacillus	968	36	41.9	912	6	ABM49392	Abm49392	Propionib
896	36	41.9	485	3	AAB29368	Aab29368	Bacillus	969	36	41.9	914	2	AAAR15785	Aaar15785	B.thuring
897	36	41.9	485	3	AAB29391	Aab29391	Bacillus	970	36	41.9	924	4	ABB67489	Abb67489	Drosophal
898	36	41.9	485	3	AAB29367	Aab29367	Bacillus	971	36	41.9	956	2	AAAR15784	Aaar15784	B.thuring
899	36	41.9	485	3	AAB29311	Aab29311	Bacillus	972	36	41.9	1100	2	AAAR15783	Aaar15783	B.thuring

973 36 41.9 1327 8 ADL98347  
 974 36 41.9 1413 4 ABB60857  
 975 36 41.9 1424 4 ABB60854  
 976 36 41.9 1510 7 ADG30698  
 977 36 41.9 1536 7 ADK64802  
 978 36 41.9 1665 7 ADF04191  
 979 36 41.9 1856 3 AAY21805  
 980 36 41.9 1856 3 AAY83274  
 981 36 41.9 2022 6 ABR63232  
 982 36 41.9 3431 8 ADI39260  
 983 36 41.9 3723 8 ADU47672  
 984 36 41.9 4970 9 AEB00354  
 985 36 41.9 36946 9 ADV97835  
 986 35.5 41.3 50 2 AAU94136  
 987 35.5 41.3 96 2 AAU94134  
 988 35.5 41.3 178 6 ABU17135  
 989 35.5 41.3 186 6 ADA33747  
 990 35.5 41.3 191 9 ADX40531  
 991 35.5 41.3 192 1 AAP61504  
 992 35.5 41.3 192 2 AAU90177  
 993 35.5 41.3 192 2 AAW89324  
 994 35.5 41.3 192 3 AAB69298  
 995 35.5 41.3 192 4 AAB85994  
 996 35.5 41.3 192 6 ABUE3326  
 997 35.5 41.3 192 6 ABUE3321  
 998 35.5 41.3 192 8 ADO26433  
 999 35.5 41.3 192 8 ADN36415  
 1000 35.5 41.3 192 9 ADX40506

## ALIGNMENTS

RESULT 1  
 ADR88216  
 ID ADR88216 standard; peptide; 14 AA.  
 AC ADR88216;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE Human heparanase epitope pep10.  
 XX  
 KW Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;  
 KW autoimmune disorder; cancer; angiogenesis; metastatic disease;  
 KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;  
 KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;  
 KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;  
 KW human; heparanase; enzyme; epitope.  
 XX  
 OS Homo sapiens.  
 XX  
 XX US2004170631-A1.  
 XX  
 PD 02-SEP-2004.  
 XX  
 PF 28-NOV-2003; 2003US-00722502.  
 XX  
 PR 02-SEP-1997; 97US-00922170.  
 PR 01-MAY-1998; 98US-00071739.  
 PR 04-NOV-1998; 98US-00186200.  
 PR 19-FEB-2003; 2003US-00368044.  
 PR 22-AUG-2003; 2003US-00645659.  
 XX  
 PA (YACO/) YACOBY-ZEEVI O.  
 PA (PERE/) PERETZ T.  
 PA (MIRO/) MIRON D.  
 PA (SHLO/) SHLOMI Y.  
 PA (PECK/) PECKER I.  
 PA (AYAL/) AYAL-HERSHKOVITZ M.  
 PA (FEIN/) FEINSTEIN E.  
 PA (VGEL/) VAN GELDER J. M.  
 PA (VLOD/) VLODAVSKY I.

PA (FRIE/) FRIEDMANN Y.  
 XX  
 PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
 PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;  
 PI Friedmann Y;  
 XX  
 XX WPI; 2004-625084/60.  
 XX  
 PT Targeted drug delivery to a heparanase-expressing tissue of a patient,  
 PT useful for treating heparanase-associated conditions such as inflammation  
 PT or cancer, comprises administering a drug and an anti-heparanase antibody  
 PT complex.  
 XX  
 PS Claim 7; SEQ ID NO 10; 58pp; English.  
 XX  
 CC The invention relates to a method of targeted drug delivery to a tissue  
 CC of a patient, the tissue expressing heparanase. The method comprises  
 CC providing a complex of a drug directly or indirectly linked to an anti-  
 CC heparanase antibody, and administering the complex to the patient. In the  
 CC targeted drug delivery, the antibody comprises an antibody or its portion  
 CC capable of specifically binding to at least one epitope of a heparanase  
 CC protein. The composition and methods of the invention are useful for  
 CC diagnosing, preventing or treating conditions associated with heparanase  
 CC catalytic activity (e.g. an inflammatory disorder, wound, scar,  
 CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell  
 CC proliferation, invasion of circulating tumour cells and metastatic  
 CC disease), for purifying heparanase, or for developing drugs for those  
 CC heparanase-associated conditions. The vasculopathy is atherosclerosis,  
 CC restenosis or aneurysm. The cancerous condition is a solid cancer or a  
 CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy  
 CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous  
 CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous  
 CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,  
 CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and  
 CC multiple myeloma. The solid cancer is selected from tumours in lip and  
 CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,  
 CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,  
 CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of  
 CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue  
 CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,  
 CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic  
 CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary  
 CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,  
 CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,  
 CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,  
 CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's  
 CC sarcoma. The present sequence is human heparanase epitope.  
 XX  
 SQ Sequence 14 AA;  
 Query Match 100.0%; Score 86; DB 8; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TWHHYILNGRTATR 14  
 |||||  
 DB 1 TWHHYILNGRTATR 14  
 RESULT 2  
 ADR78183  
 ID ADR78183 standard; peptide; 14 AA.  
 XX  
 AC ADR78183;  
 XX  
 XX 13-JAN-2005 (first entry)  
 DT  
 XX  
 DE Functional peptide epitope of human heparanase, pep10.  
 XX  
 KW Antibody; epitope; heparanase; pathological condition; angiogenesis;  
 KW cell proliferation; cancerous condition; tumour cell invasion; disorder;  
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;  
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;

KW cytotatic; antiinflammatory; vulnery; antiarteriosclerotic;  
 KW vasotrophic; immunosuppressive; nephrotropic; antidiabetic; human.  
 OS Homo sapiens.

XX US2004213789-A1.  
 XX 28-OCT-2004.

XX 22-AUG-2003; 2003US-00645659.

XX 02-SEP-1997; 97US-00922170.

PR 01-MAY-1998; 98US-00071739.

PR 04-NOV-1998; 98US-00186200.

PR 19-FEB-2003; 2003US-00368044.

XX (YACO/) YACOBY-ZEEVI O.

PA (PERE/) PERETZ T.

PA (MIRO/) MIRON D.

PA (SHLO/) SHLOMI Y.

PA (PECK/) PECKER I.

PA (AYAL/) AYAL-HERSHKOVITZ M.

PA (FEIN/) FEINSTEIN E.

PA (GELD/) GELDER J M V.

PA (VLOD/) VLADAVSKI I.

PA (FRIE/) FRIEDMANN Y.

XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;

PI Ayal-Hershkovitz M, Feinstein E, Gelder JMV, Vlodavsky I;

PI Friedmann Y;

XX WPI; 2004-774790/76.

XX New neutralizing monoclonal anti-heparanase antibodies, useful for  
 PT detecting, treating or preventing cancer, inflammatory or autoimmune  
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.

XX Claim 67; SEQ ID NO 10; 69pp; English.

XX The invention relates to an isolated antibody or antibody portion capable  
 CC of specifically binding to or elicited by at least one epitope of a  
 CC heparanase protein, where the heparanase protein is at least 60%  
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and  
 CC where at least one epitope comprises a sequence at least 70% homologous  
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)  
 CC a hybridoma cell line comprising a cell line for producing the monoclonal  
 CC antibody, (b) a method for detecting, treating or preventing a  
 CC pathological condition or a heparanase-related disorder or condition in a  
 CC subject, (c) a method for monitoring the state of a heparanase-related  
 CC disorder or condition in a subject, and (d) a pharmaceutical composition  
 CC comprising the isolated anti-heparanase antibody or antibody portion and  
 CC a pharmaceutical carrier. The antibody, methods, and composition are  
 CC useful for detecting, treating, preventing or monitoring a pathological  
 CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition  
 CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,  
 CC or prostate cancer), minor cell proliferation, invasion of circulating  
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder  
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy  
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or  
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,  
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell  
 CC carcinoma) in a mammal. This sequence represents a functional peptide  
 CC epitope of human heparanase.

XX Sequence 14 AA;

Query Match 100.0%; Score 86; DB 8; Length 14;

Best Local Similarity 100.0%; Pred. No. 1.4e-07; Gaps 0;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14

DB 1 TWHYYLNGRTATR 14

RESULT 3  
 ID AEA42432

XX AEA42432 standard; peptide; 14 AA.

XX AEA42432;

XX 28-JUL-2005 (first entry)

XX Human heparanase epitope peptide SEQ ID NO:10.

XX antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;  
 KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotrophic;  
 KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;  
 KW angiogenesis disorder; cancer; tumor; metastasis; epitope.

OS Homo sapiens.

XX AU2004201462-A1.

XX 06-MAY-2004.

XX 08-APR-2004; 2004AU-00201462.

XX 08-APR-2004; 2004AU-00201462.

XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;

PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;

PI Feinstein E;

XX WPI; 2005-173343/19.

XX Novel isolated antibody capable of specifically binding to epitope of  
 PT heparanase protein, useful for preventing and treating heparanase-related  
 PT disorder such as inflammatory disorder, scars, autoimmune conditions or  
 PT angiogenesis.

XX Claim 7; SEQ ID NO 10; 260pp; English.

XX The invention relates to an isolated antibody or its portion (I) capable  
 CC of specifically binding to an epitope of a heparanase protein. Also  
 CC described: (1) a cell line (II) for producing a monoclonal antibody or  
 CC its portion, comprising a cell line for producing (I); (2) a  
 CC pharmaceutical composition comprising (I) and a carrier; and (3) an  
 CC affinity medium (III) for binding human heparanase polypeptides,  
 CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)  
 CC useful for treating a subject suffering from a pathological condition,  
 CC which involves administering (I) to the subject. (I) is useful for  
 CC preventing and treating heparanase-related disorder or condition chosen  
 CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune  
 CC condition, angiogenesis, cell proliferation, cancerous condition, tumor  
 CC cell proliferation, invasion of circulating tumor cells and metastatic  
 CC disease. (I) is useful for detecting the presence of heparanase-  
 CC polypeptide in a sample. (I) is useful for detecting heparanase-related  
 CC disease or condition in a subject such as vertebrate, preferably mammal  
 CC e.g., human. The heparanase-related disorder or condition further  
 CC includes renal disease or disorder chosen from diabetic nephropathy,  
 CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome  
 CC and renal cell carcinoma. The present sequence represents a human  
 CC heparanase epitope peptide, which is used in the exemplification of the  
 CC present invention.

XX Sequence 14 AA;

Query Match 100.0%; Score 86; DB 9; Length 14;

Best Local Similarity 100.0%; Pred. No. 1.4e-07; Gaps 0;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14



```

Db      1 TWHHYLLNGRTATR 14
|||||
RESULT 4
ADU71059
ID ADU71059 standard; peptide; 15 AA.
XX
AC ADU71059;
XX
DT 10-FEB-2005 (first entry)
XX
DE Human heparanase peptide SEQ ID NO:744.
XX
KW enzyme; heparinase; vaccine; human leukocyte antigen; HLA;
KW immunostimulant; cytostatic; immune disorder; metastasis.
XX
OS Homo sapiens.
XX
FN EP1479764-A1.
XX
PD 24-NOV-2004.
XX
PF 19-MAY-2003; 2003EP-00011038.
XX
PR 19-MAY-2003; 2003EP-00011038.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA (UYHE-) UNIV HEIDELBERG RUPRECHT-KARLS.
XX
PI Schirmacher V, Beckhove P, Sommerfeldt N;
XX
WPI; 2005-014847/02.
XX
New heparanase nonapeptide that binds to a human leukocyte antigen (HLA)
molecule or its functional derivative, useful for preparing a medicament
for inducing an immune response or for treating metastatic tumors.
XX
Homo sapiens.
XX
EP1479764-A1.
XX
PD 24-NOV-2004.
XX
PF 19-MAY-2003; 2003EP-00011038.
XX
PR 19-MAY-2003; 2003EP-00011038.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA (UYHE-) UNIV HEIDELBERG RUPRECHT-KARLS.
XX
PI Schirmacher V, Beckhove P, Sommerfeldt N;
XX
WPI; 2005-014847/02.
XX
New heparanase nonapeptide that binds to a human leukocyte antigen (HLA)
molecule or its functional derivative, useful for preparing a medicament
for inducing an immune response or for treating metastatic tumors.
XX
Disclosure; SEQ ID NO 744; 269pp; English.
XX
The invention relates to a novel heparanase peptide that binds to a human
leukocyte antigen (HLA) molecule, where the peptide is a nonapeptide, or
its functional derivative. A peptide of the invention has immunostimulant
and cytostatic activity, and is used in a vaccine. The heparinase peptide
is useful for preparing a medicament which induces an immune response or
for treating metastatic tumors. The present sequence represents a
heparinase peptide of the invention.
XX
Sequence 15 AA;
Query Match 100.0%; Score 86; DB 9; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TWHHYLLNGRTATR 14
|||||
Db 2 TWHHYLLNGRTATR 15
|||||
RESULT 5
ADU71060
ID ADU71060 standard; peptide; 15 AA.
XX
AC ADU71060;
XX
DT 10-FEB-2005 (first entry)
XX
DE Human heparanase peptide SEQ ID NO:745.
XX
KW enzyme; heparinase; vaccine; human leukocyte antigen; HLA;
KW immunostimulant; cytostatic; immune disorder; metastasis.
XX
OS Homo sapiens.

```

```

XX
FN EP1479764-A1.
XX
PD 24-NOV-2004.
XX
PF 19-MAY-2003; 2003EP-00011038.
XX
PR 19-MAY-2003; 2003EP-00011038.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA (UYHE-) UNIV HEIDELBERG RUPRECHT-KARLS.
XX
PI Schirmacher V, Beckhove P, Sommerfeldt N;
XX
WPI; 2005-014847/02.
XX
New heparanase nonapeptide that binds to a human leukocyte antigen (HLA)
molecule or its functional derivative, useful for preparing a medicament
for inducing an immune response or for treating metastatic tumors.
XX
Disclosure; SEQ ID NO 745; 269pp; English.
XX
The invention relates to a novel heparanase peptide that binds to a human
leukocyte antigen (HLA) molecule, where the peptide is a nonapeptide, or
its functional derivative. A peptide of the invention has immunostimulant
and cytostatic activity, and is used in a vaccine. The heparinase peptide
is useful for preparing a medicament which induces an immune response or
for treating metastatic tumors. The present sequence represents a
heparinase peptide of the invention.
XX
Sequence 15 AA;
Query Match 100.0%; Score 86; DB 9; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TWHHYLLNGRTATR 14
|||||
Db 1 TWHHYLLNGRTATR 14
|||||
RESULT 6
ADU27047
ID ADU27047 standard; peptide; 15 AA.
XX
AC ADU27047;
XX
DT 05-MAY-2005 (first entry)
XX
DE Heparanase inhibitor peptide #6.
XX
KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
KW neurological disease; viral infection; infection; cytostatic;
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
KW heparanase modulator; enzyme purification.
XX
OS Synthetic.
XX
FN WO2005016227-A2.
XX
PD 24-FEB-2005.
XX
PF 12-AUG-2004; 2004WO-IL000744.
XX
PR 14-AUG-2003; 2003US-0494800P.
PR 12-JAN-2004; 2004US-0535492P.
XX
PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
XX
PI Van-Gelder JM, Miron D;
XX
WPI; 2005-182203/19.
XX

```

PT Regulating heparanase activity, useful for treating heparanase-associated  
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases, heparanase  
 PT activation.

XX Claim 56; SEQ ID NO 21; 211pp; English.

XX The invention relates to a method of regulating heparanase activity in a  
 CC tissue or regulating a biological process depending at least in part on  
 CC heparanase activity comprising modulating heparanase activation. The  
 CC invention also relates to methods of treating a heparanase- or heparin  
 CC binding protein-associated disease or disorder in a subject, a  
 CC pharmaceutical composition for use in the treatment of a heparanase-  
 CC associated disease or disorder comprising a therapeutic amount of an  
 CC agent capable of modulating heparanase activation and a pharmaceutical  
 CC carrier or diluent, a method of identifying a protease activator of  
 CC heparanase, a protease substrate mimetic comprising a peptide  
 CC representing a subset or all substrate residues or cleavage sites of  
 CC human heparanase or an equivalent non-human heparanase, a method of  
 CC producing active heparanase and a method of modulating an adhesion  
 CC activity of heparanase. The composition and methods are useful for  
 CC modulating heparanase activation and for treating heparanase-associated  
 CC diseases or disorders such as cancer, inflammation, cardiovascular  
 CC diseases, neurological diseases or viral infections. This sequence  
 CC represents a heparanase inhibitor peptide used in the scope of the  
 CC invention.

XX Sequence 15 AA;

Query Match 100.0%; Score 86; DB 9; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHHYLLNGRTATR 14  
 Db |||||

RESULT 7

ID ADR88207 standard; protein; 386 AA.

XX AC ADR88207;

XX DT 18-NOV-2004 (first entry)

XX DE Human mature heparanase 45 kDa major subunit.

XX Targeted drug delivery ; inflammatory disorder; wound; scar;  
 KW vasculopathy; autoimmune disorder; cancer; angiogenesis;  
 KW metastatic disease; atherosclerosis; reestenosis; aneurysm; solid cancer;  
 KW non-solid cancer; haematopoietic malignancy ; lymphocytic leukaemia;  
 KW myelogenous leukaemia; Hodgkin's disease; multiple myeloma;  
 KW haemangiosarcoma; Kaposi's sarcoma; human ; heparanase; enzyme.

XX OS Homo sapiens.

XX PN US2004170631-A1.

XX PD 02-SEP-2004.

XX PF 28-NOV-2003; 2003US-00722502.

XX PR 02-SEP-1997; 97US-00922170.

XX PR 01-MAY-1998; 98US-00071739.

XX PR 04-NOV-1998; 98US-00186200.

XX PR 19-FEB-2003; 2003US-00368044.

XX PR 22-AUG-2003; 2003US-00645659.

XX (YACO/) YACOBY-ZEEVI O.

PA (PERE/) PERETZ T.

PA (MIRO/) MIRON D.

PA (SHLO/) SHLOMI Y.

PA (PECK/) PECKER I.  
 PA (AYAL/) AYAL-HERSHKOVITZ M.  
 PA (FEIN/) FEINSTEIN E.  
 PA (VGEL/) VAN GELDER J M.  
 PA (VLOD/) VLODAVSKY I.  
 PA (FRIE/) FRIEDMANN Y.

XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
 PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;  
 PI Friedmann Y;

XX WPI; 2004-625084/60.

XX Targeted drug delivery to a heparanase-expressing tissue of a patient,  
 PT useful for treating heparanase-associated conditions such as inflammation  
 PT or cancer, comprises administering a drug and an anti-heparanase antibody  
 PT complex.

XX Claim 2; SEQ ID NO 1; 58pp; English.

XX The invention relates to a method of targeted drug delivery to a tissue  
 CC of a patient, the tissue expressing heparanase. The method comprises  
 CC providing a complex of a drug directly or indirectly linked to an anti-  
 CC heparanase antibody, and administering the complex to the patient. In the  
 CC targeted drug delivery, the antibody comprises an antibody or its portion  
 CC capable of specifically binding to at least one epitope of a heparanase  
 CC protein. The composition and methods of the invention are useful for  
 CC diagnosing, preventing or treating conditions associated with heparanase  
 CC catalytic activity (e.g. an inflammatory disorder, wound, scar,  
 CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell  
 CC proliferation, invasion of circulating tumour cells and metastatic  
 CC disease), for purifying heparanase, or for developing drugs for those  
 CC heparanase-associated conditions. The vasculopathy is atherosclerosis,  
 CC reestenosis or aneurysm. The cancerous condition is a solid cancer or a  
 CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy  
 CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous  
 CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous  
 CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,  
 CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and  
 CC multiple myeloma. The solid cancer is selected from tumours in lip and  
 CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,  
 CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,  
 CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of  
 CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue  
 CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,  
 CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic  
 CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary  
 CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,  
 CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,  
 CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,  
 CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's  
 CC sarcoma. The present sequence is the 45 kDa major subunit of human mature  
 CC heparanase.

XX SQ Sequence 386 AA;

Query Match 100.0%; Score 86; DB 8; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 6e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHHYLLNGRTATR 14  
 Db |||||

RESULT 8  
 ADR78174

ID ADR78174 standard; protein; 386 AA.

XX AC ADR78174;

XX DT 13-JAN-2005 (first entry)

DE 45kDa subunit of mature processed human heparanase dimer.

XX Antibody; epitope; heparanase; pathological condition; angiogenesis;  
KW cell proliferation; cancerous condition; tumour cell invasion;  
KW metastatic disease; heparanase-related disorder; inflammatory disorder;  
KW wound; scar; vasculopathy; autoimmune condition; renal disease;  
KW cytostatic; antiinflammatory; vulnerable; antiarteriosclerotic;  
KW vasotrophic; immunosuppressive; nephrotropic; antidiabetic; human.  
XX Homo sapiens.

OS  
XX  
XX US2004213789-A1.  
XX  
XX 28-OCT-2004.  
XX  
XX 22-AUG-2003; 2003US-00645659.  
XX  
XX 02-SEP-1997; 97US-00922170.  
PR 01-MAY-1998; 98US-00071739.  
PR 04-NOV-1998; 98US-00186200.  
PR 19-FEB-2003; 2003US-00368044.  
XX (YACO/) YACOBY-ZEEVI O.  
PA (PERE/) PERETZ T.  
PA (MIRO/) MIRON D.  
PA (SHLO/) SHLOMI Y.  
PA (PECK/) PECKER I.  
PA (AYAL/) AYAL-HERSHKOVITZ M.  
PA (FEIN/) FEINSTEIN E.  
PA (GELD/) GELDER J M V.  
PA (VLOD/) VLODAVSKY I.  
PA (FRIE/) FRIEDMANN Y.  
XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;  
PI Friedmann Y;  
XX WPI; 2004-774790/76.  
XX  
XX New neutralizing monoclonal anti-heparanase antibodies, useful for  
PT detecting, treating or preventing cancer, inflammatory or autoimmune  
PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.  
XX  
XX Claim 5; SEQ ID NO 1; 68pp; English.

XX The invention relates to an isolated antibody or antibody portion capable  
CC of specifically binding to or elicited by at least one epitope of a  
CC heparanase protein, where the heparanase protein is at least 60%  
CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and  
CC where at least one epitope comprises a sequence at least 70% homologous  
CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)  
CC a hybridoma cell line comprising a cell line for producing the monoclonal  
CC antibody, (b) a method for detecting, treating or preventing a  
CC pathological condition or a heparanase-related disorder or condition in a  
CC subject, (c) a method for monitoring the state of a heparanase-related  
CC disorder or condition in a subject, and (d) a pharmaceutical composition  
CC comprising the isolated anti-heparanase antibody or antibody portion and  
CC a pharmaceutical carrier. The antibody, methods, and composition are  
CC useful for detecting, treating, preventing or monitoring a pathological  
CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition  
CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,  
CC or prostate cancer), minor cell proliferation, invasion of circulating  
CC tumour cells, or a metastatic disease, or a heparanase-related disorder  
CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy  
CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or  
CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,  
CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell  
CC carcinoma) in a mammal. This sequence represents the 45kDa subunit of  
CC mature processed human heparanase dimer.

XX  
XX Sequence 386 AA;  
SQ Query Match 100.0%; Score 86; DB 8; Length 386;

Best Local Similarity 100.0%; Pred. No. 6e-06; Indels 0; Gaps 0;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14  
DB 137 TWHYYLNGRTATR 150

RESULT 9  
ADY27057  
ID ADY27057 standard; protein; 386 AA.  
XX  
AC ADY27057;  
XX  
DT 05-MAY-2005 (first entry)  
XX  
DE Heparanase inhibitor protein #1.  
XX  
XX Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
KW neurological disease; viral infection; infection; cytostatic;  
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
KW heparanase modulator; enzyme purification.  
XX  
OS Homo sapiens.  
XX  
XX WO2005016227-A2.  
XX  
XX 24-FEB-2005.  
XX  
XX 12-AUG-2004; 2004WO-IL000744.  
XX  
XX 14-AUG-2003; 2003US-0494800P.  
PR 12-JAN-2004; 2004US-0535492P.  
XX  
XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.  
XX  
XX Van-Gelder JM, Miron D;  
XX  
XX WPI; 2005-182203/19.  
XX  
XX Regulating heparanase activity, useful for treating heparanase-associated  
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
PT neurological diseases or viral diseases) comprises modulating heparanase  
PT activation.  
XX  
XX Claim 55; SEQ ID NO 33; 211pp; English.

XX The invention relates to a method of regulating heparanase activity in a  
CC tissue or regulating a biological process depending at least in part on  
CC heparanase activity comprising modulating heparanase activation. The  
CC invention also relates to methods of treating a heparanase- or heparin  
CC binding protein-associated disease or disorder in a subject, a  
CC pharmaceutical composition for use in the treatment of a heparanase-  
CC associated disease or disorder comprising a therapeutic amount of an  
CC agent capable of modulating heparanase activation and a pharmaceutical  
CC carrier or diluent, a method of identifying a protease activator of  
CC heparanase, a protease substrate mimetic comprising a peptide  
CC representing a subset or all substrate residues or cleavage sites of  
CC human heparanase or an equivalent non-human heparanase, a method of  
CC producing active heparanase and a method of modulating an adhesion  
CC activity of heparanase. The composition and methods are useful for  
CC modulating heparanase activation and for treating heparanase-associated  
CC diseases or disorders such as cancer, inflammation, cardiovascular  
CC diseases, neurological diseases or viral infections. This sequence  
CC represents a heparanase inhibitor protein used in the scope of the  
CC invention.  
XX  
XX Sequence 386 AA;  
SQ Query Match 100.0%; Score 86; DB 9; Length 386;  
Best Local Similarity 100.0%; Pred. No. 6e-06; Indels 0; Gaps 0;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14  
 |||||  
 Db 137 TWHYYLNGRTATR 150

RESULT 10  
 ADZ18995  
 ID ADZ18995 standard; protein; 386 AA.  
 AC ADZ18995;  
 XX  
 XX 16-JUN-2005 (first entry)  
 XX  
 XX Human heparanase consensus cleavage site #2.  
 XX  
 KW Enzyme engineering; heparanase; metastasis; autoimmune disease;  
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
 KW immunosuppressive; enzyme.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2005030962-A1.  
 XX  
 XX 07-APR-2005.  
 XX  
 PF 17-SEP-2004; 2004WO-EP010517.  
 XX  
 PR 26-SEP-2003; 2003US-0506479P.  
 PR 20-JAN-2004; 2004US-0537729P.  
 XX  
 XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.  
 PA  
 XX Lahm A, Nardella C, Pallaoaro M, Steinkuhler C;  
 XX WPI; 2005-273382/28.  
 DR  
 XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.  
 XX  
 PS Disclosure; SEQ ID NO 16; 65pp; English.  
 XX  
 CC The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoproteinase. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in  
 CC inhibitor screening assays for the development of therapeutics or  
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a human heparanase  
 CC consensus cleavage site used in the scope of the invention.  
 XX  
 SQ Sequence 386 AA;

Query Match 100.0%; Score 86; DB 9; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 6e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14  
 |||||  
 Db 137 TWHYYLNGRTATR 150

RESULT 11  
 AEA42423  
 ID AEA42423 standard; protein; 386 AA.  
 AC AEA42423;  
 XX  
 XX 28-JUL-2005 (first entry)  
 DT  
 XX Human mature heparanase dimer 45 kDa subunit SEQ ID NO:1.  
 DE  
 XX

KW antibody; heparanase; antiinflammatory; vulnary; immunosuppressive;  
 KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;  
 KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;  
 KW angiogenesis disorder; cancer; tumor; metastasis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN AU2004201462-A1.  
 XX  
 XX 06-MAY-2004.  
 XX  
 PF 08-APR-2004; 2004AU-00201462.  
 XX  
 XX 08-APR-2004; 2004AU-00201462.  
 PR  
 XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.  
 XX  
 XX Vlodayevy I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;  
 PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;  
 PI Feinstein E;  
 XX  
 DR WPI; 2005-173343/19.  
 XX  
 XX Novel isolated antibody capable of specifically binding to epitope of  
 PT heparanase protein, useful for preventing and treating heparanase-related  
 PT disorder such as inflammatory disorder, scars, autoimmune conditions or  
 PT angiogenesis.  
 XX  
 PS Claim 2; SEQ ID NO 1; 260pp; English.  
 XX  
 CC The invention relates to an isolated antibody or its portion (I) capable  
 CC of specifically binding to an epitope of a heparanase protein. Also  
 CC described: (1) a cell line (II) for producing a monoclonal antibody or  
 CC its portion, comprising a cell line for producing (I); (2) a  
 CC pharmaceutical composition comprising (I) and a carrier; and (3) an  
 CC affinity medium (III) for binding human heparanase polypeptides,  
 CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)  
 CC useful for treating a subject suffering from a pathological condition,  
 CC which involves administering (I) to the subject. (I) is useful for  
 CC preventing and treating heparanase-related disorder or condition chosen  
 CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune  
 CC condition, angiogenesis, cell proliferation, cancerous condition, tumor  
 CC cell proliferation, invasion of circulating tumor cells and metastatic  
 CC disease. (I) is useful for detecting the presence of heparanase  
 CC polypeptide in a sample. (I) is useful for detecting heparanase-related  
 CC disease or condition in a subject such as vertebrate, preferably mammal  
 CC e.g., human. The heparanase-related disorder or condition further  
 CC includes renal disease or disorder chosen from diabetic nephropathy,  
 CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome  
 CC and renal cell carcinoma. The present sequence represents the 45 kDa  
 CC subunit of the human mature processed heparanase dimer, which is used in  
 CC the exemplification of the present invention.  
 XX  
 SQ Sequence 386 AA;

Query Match 100.0%; Score 86; DB 9; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 6e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14  
 |||||  
 Db 137 TWHYYLNGRTATR 150

RESULT 12  
 ADY27061  
 ID ADY27061 standard; protein; 460 AA.  
 XX  
 AC ADY27061;  
 XX  
 XX 05-MAY-2005 (first entry)  
 DT  
 XX

DE Heparanase inhibitor protein #4.  
 KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
 KW neurological disease; viral infection; infection; cytostatic;  
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
 KW heparanase modulator; enzyme purification.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2005016227-A2.  
 XX  
 PD 24-FEB-2005.  
 XX  
 PF 12-AUG-2004; 2004WO-IL000744.  
 XX  
 PR 14-AUG-2003; 2003US-0494800P.  
 PR 12-JAN-2004; 2004US-0535492P.  
 XX  
 PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.  
 XX  
 PI Van-Gelder JM, Miron D;  
 XX  
 DR WPI; 2005-182203/19.  
 XX  
 PT Regulating heparanase activity, useful for treating heparanase-associated  
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
 PT neurological diseases or viral diseases) comprises modulating heparanase  
 PT activation.  
 XX  
 PS Disclosure; SEQ ID NO 37; 211pp; English.  
 XX  
 CC The invention relates to a method of regulating heparanase activity in a  
 CC tissue or regulating a biological process depending at least in part on  
 CC heparanase activity comprising modulating heparanase activation. The  
 CC invention also relates to methods of treating a heparanase- or heparin  
 CC binding protein-associated disease or disorder in a subject, a  
 CC pharmaceutical composition for use in the treatment of a heparanase-  
 CC associated disease or disorder comprising a therapeutic amount of an  
 CC agent capable of modulating heparanase activation and a pharmaceutical  
 CC carrier or diluent, a method of identifying a protease activator of  
 CC heparanase, a protease substrate mimetic comprising a peptide  
 CC representing a subset or all substrate residues or cleavage sites of  
 CC human heparanase or an equivalent non-human heparanase, a method of  
 CC producing active heparanase and a method of modulating an adhesion  
 CC activity of heparanase. The composition and methods are useful for  
 CC modulating heparanase activation and for treating heparanase-associated  
 CC diseases or disorders such as cancer, inflammation, cardiovascular  
 CC diseases, neurological diseases or viral infections. This sequence  
 CC represents a heparanase inhibitor protein used in the scope of the  
 CC invention.  
 XX  
 SQ Sequence 460 AA;  
 Query Match 100.0%; Score 86; DB 9; Length 460;  
 Best Local Similarity 100.0%; Pred. No. 7.3e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TWHYYLNGRTATR 14  
 |||||  
 DB 211 TWHYYLNGRTATR 224  
 RESULT 13  
 AEB87589  
 ID AEB87589 standard; protein; 486 AA.  
 XX  
 AC AEB87589;  
 XX  
 DT 06-OCT-2005 (first entry)  
 XX  
 DE Human heparanase 65delta20 deletion mutant.  
 XX  
 KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;

KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;  
 KW leukemia; lymphoma; dermatological disease; hematological disease;  
 KW immune disorder; inflammation; antiinflammatory; renal disease;  
 KW nephrotropic; endocrine disease; genitourinary disease;  
 KW autoimmune disease; immunosuppressive; drug screening; mutein.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO2005071070-A2.  
 XX  
 PD 04-AUG-2005.  
 XX  
 PF 20-JAN-2005; 2005WO-IL000068.  
 XX  
 PR 22-JAN-2004; 2004IL-00160025.  
 PR 28-JUL-2004; 2004US-00901943.  
 XX  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.  
 XX  
 PI Vlodavsky I, Ilan N, Levy-Adam F;  
 XX  
 DR WPI; 2005-564219/57.  
 DR N-PSDB; AEB87588.  
 XX  
 PT New amino acid sequences derived from the 50 kDa subunit of heparanase,  
 PT for treating or inhibiting malignant proliferative, inflammatory, kidney  
 PT disorder or autoimmune disorder.  
 XX  
 PS Claim 107; SEQ ID NO 31; 167pp; English.  
 XX  
 CC The present sequence is that of a deletion mutant of human heparanase,  
 CC denoted 65delta20, which is devoid of amino acid residues 411-432 of the  
 CC native protein. The recombinant protein is deficient of heparanase  
 CC endoglycosidase catalytic activity. The invention relates to amino acid  
 CC sequences derived from the N-terminus region of the 50 kDa subunit of  
 CC heparanase, particularly in the regions between amino acid residues 158-  
 CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise  
 CC heparin-binding domains. The invention also provides an antibody directed  
 CC to these sequences, in particular the 158-171 peptide, and compositions  
 CC and uses of this antibody as a heparanase inhibitor. A screening method  
 CC is provided for specific heparanase inhibitors. Claimed pharmaceutical  
 CC compositions comprising (i) a peptide derived from the N-terminus region  
 CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a  
 CC peptide, or (iii) an antibody which specifically recognizes the peptide  
 CC are used for the inhibition of heparanase catalytic activity associated  
 CC with an inflammatory disorder, kidney disease, autoimmune disease,  
 CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or  
 CC with a malignant proliferative disorder, especially a solid or non-solid  
 CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.  
 XX  
 SQ Sequence 486 AA;  
 Query Match 100.0%; Score 86; DB 9; Length 486;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TWHYYLNGRTATR 14  
 |||||  
 DB 259 TWHYYLNGRTATR 272  
 RESULT 14  
 ADZ18996  
 ID ADZ18996 standard; protein; 492 AA.  
 XX  
 AC ADZ18996;  
 XX  
 DT 16-JUN-2005 (first entry)  
 XX  
 DE Hep106 construct protein.  
 XX  
 KW Enzyme engineering; heparanase; hep106; metastasis; autoimmune disease;

KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
 KW immunosuppressive; enzyme.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2005030962-A1.  
 PD 07-APR-2005.  
 XX  
 PF 17-SEP-2004; 2004WO-EP010517.  
 XX  
 PR 26-SEP-2003; 2003US-0506479P.  
 PR 20-JAN-2004; 2004US-0537729P.  
 XX  
 PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.  
 XX  
 PI Lahm A, Nardella C, Pallao M, Steinkuhler C;  
 XX  
 DR WPI; 2005-273382/28.  
 DR N-PSDB; ADZ18997.  
 XX  
 PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.  
 XX  
 PS Example 2; SEQ ID NO 17; 65pp; English.  
 XX  
 CC The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoproteinase. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in  
 CC inhibitor screening assays for the development of therapeutics or  
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a hepl06 construct protein  
 CC used in the scope of the invention.  
 XX  
 SQ Sequence 492 AA;  
 Query Match 100.0%; Score 86; DB 9; Length 492;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TWHYYLNGRTATR 14  
 |||||  
 Db 243 TWHYYLNGRTATR 256  
 |||||  
 RESULT 15  
 AEB87562  
 ID AEB87562 standard; protein; 493 AA.  
 XX  
 AC AEB87562;  
 XX  
 DT 06-OCT-2005 (first entry)  
 XX  
 DE Human heparanase 65delta15 deletion mutant.  
 XX  
 KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;  
 KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;  
 KW leukemia; lymphoma; dermatological disease; hematological disease;  
 KW immune disorder; inflammation; antiinflammatory; renal disease;  
 KW nephrotropic; endocrine disease; genitourinary disease;  
 KW autoimmune disease; immunosuppressive; drug screening; mutein.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO2005071070-A2.  
 XX  
 PD 04-AUG-2005.  
 XX  
 PF 20-JAN-2005; 2005WO-IL000068.

XX 22-JAN-2004; 2004IL-00160025.  
 PR 28-JUL-2004; 2004US-00901943.  
 XX  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.  
 XX  
 PI Vlodevsky I, Ilan N, Levy-Adam F;  
 XX  
 DR WPI; 2005-564219/57.  
 DR N-PSDB; AEB87561.  
 XX  
 PT New amino acid sequences derived from the 50 kDa subunit of heparanase,  
 PT for treating or inhibiting malignant proliferative, inflammatory, kidney  
 PT disorder or autoimmune disorder.  
 XX  
 PS Claim 105; SEQ ID NO 4; 167pp; English.  
 XX  
 CC The present sequence is that of a deletion mutant of human heparanase,  
 CC denoted 65delta15, which is devoid of amino acid residues 158-171 of the  
 CC native protein. The recombinant protein is deficient of heparanase  
 CC endoglycosidase catalytic activity. The invention relates to amino acid  
 CC sequences derived from the N-terminus region of the 50 kDa subunit of  
 CC heparanase, particularly in the regions between amino acid residues 158-  
 CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise  
 CC heparin-binding domains. The invention also provides an antibody directed  
 CC to these sequences, in particular the 158-171 peptide, and compositions  
 CC and uses of this antibody as a heparanase inhibitor. A screening method  
 CC is provided for specific heparanase inhibitors. Claimed pharmaceutical  
 CC compositions comprising (i) a peptide derived from the N-terminus region  
 CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a  
 CC peptide, or (iii) an antibody which specifically recognizes the peptide  
 CC are used for the inhibition of heparanase catalytic activity associated  
 CC with an inflammatory disorder, kidney disease, autoimmune disease,  
 CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or  
 CC with a malignant proliferative disorder, especially a solid or non-solid  
 CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.  
 XX  
 SQ Sequence 493 AA;  
 Query Match 100.0%; Score 86; DB 9; Length 493;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TWHYYLNGRTATR 14  
 |||||  
 Db 244 TWHYYLNGRTATR 257  
 |||||  
 RESULT 16  
 ADZ18999  
 ID ADZ18999 standard; protein; 495 AA.  
 XX  
 AC ADZ18999;  
 XX  
 DT 16-JUN-2005 (first entry)  
 XX  
 DE Hep109 construct protein.  
 XX  
 KW Enzyme engineering; heparanase; hepl09; metastasis; autoimmune disease;  
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
 KW immunosuppressive; enzyme.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2005030962-A1.  
 XX  
 PD 07-APR-2005.  
 XX  
 PF 17-SEP-2004; 2004WO-EP010517.  
 XX  
 PR 26-SEP-2003; 2003US-0506479P.  
 PR 20-JAN-2004; 2004US-0537729P.  
 XX

PA (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.  
XX Lahm A, Nardella C, Pallaoro M, Steinkuhler C;  
XX WPI; 2005-273382/28.  
DR N-PSDB; ADZ18998.  
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
PT nucleotide sequence that encodes mammalian heparanase protein and has two  
PT consensus cleavage sites located between specific nucleotide encoding  
PT residues.  
XX Example 2; SEQ ID NO 20; 65pp; English.  
XX The invention relates to a synthetic nucleic acid molecule that encodes  
CC mammalian heparanase protein, where the nucleic acid comprises two  
CC consensus cleavage sites recognized by endoproteinase. The sequences are  
CC useful for expressing mammalian heparanase in non-mammalian cells and in  
CC inhibitor screening assays for the development of therapeutics or  
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
CC and/or inflammation. This sequence represents a hepi09 construct protein  
CC used in the scope of the invention.  
XX Sequence 495 AA;  
SQ  
Query Match 100.0%; Score 86; DB 9; Length 495;  
Best Local Similarity 100.0%; Pred. No. 7.9e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TWHYYLNGRTATR 14  
DB 246 TWHYYLNGRTATR 259  
|||||  
RESULT 17  
AEB87587  
ID AEB87587 standard; protein; 497 AA.  
XX AC AEB87587;  
XX 06-OCT-2005 (first entry)  
XX Human heparanase 65delta10 deletion mutant.  
XX Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;  
KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;  
KW leukemia; lymphoma; dermatological disease; hematological disease;  
KW immune disorder; inflammation; antiinflammatory; renal disease;  
KW nephrotropic; endocrine disease; genitourinary disease;  
KW autoimmune disease; immunosuppressive; drug screening; mutein.  
XX Homo sapiens.  
OS Synthetic.  
XX WO2005071070-A2.  
XX 04-AUG-2005.  
XX 20-JAN-2005; 2005WO-IL000068.  
XX 22-JAN-2004; 2004IL-00160025.  
PR 28-JUL-2004; 2004US-00901943.  
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.  
PA Vlodavsky I, Ilan N, Levy-Adam F;  
XX WPI; 2005-564219/57.  
DR N-PSDB; AEB87586.  
XX New amino acid sequences derived from the 50 kDa subunit of heparanase,  
PT for treating or inhibiting malignant proliferative, inflammatory, kidney  
PT disorder or autoimmune disorder.

XX Claim 106; SEQ ID NO 29; 167pp; English.  
XX The present sequence is that of a deletion mutant of human heparanase,  
CC denoted 65delta10, which is devoid of amino acid residues 270-280 of the  
CC native protein. The recombinant protein is deficient of heparanase  
CC endoglycosidase catalytic activity. The invention relates to amino acid  
CC sequences derived from the N-terminus region of the 50 kDa subunit of  
CC heparanase, particularly in the regions between amino acid residues 158-  
CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise  
CC heparin-binding domains. The invention also provides an antibody directed  
CC to these sequences, in particular the 158-171 peptide, and compositions  
CC and uses of this antibody as a heparanase inhibitor. A screening method  
CC is provided for specific heparanase inhibitors. Claimed pharmaceutical  
CC compositions comprising (i) a peptide derived from the N-terminus region  
CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a  
CC peptide, or (iii) an antibody which specifically recognizes the peptide  
CC are used for the inhibition of heparanase catalytic activity associated  
CC with an inflammatory disorder, kidney disease, autoimmune disease,  
CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or  
CC with a malignant proliferative disorder, especially a solid or non-solid  
CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.  
XX Sequence 497 AA;  
SQ  
Query Match 100.0%; Score 86; DB 9; Length 497;  
Best Local Similarity 100.0%; Pred. No. 8e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TWHYYLNGRTATR 14  
DB 248 TWHYYLNGRTATR 261  
|||||  
RESULT 18  
ADZ19000  
ID ADZ19000 standard; protein; 501 AA.  
XX AC ADZ19000;  
XX 16-JUN-2005 (first entry)  
XX HepGS3 construct protein.  
XX Enzyme engineering; heparanase; hepGS3; metastasis; autoimmune disease;  
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
KW immunosuppressive; enzyme.  
XX Synthetic.  
XX WO2005030962-A1.  
XX 07-APR-2005.  
XX 17-SEP-2004; 2004WO-EP010517.  
XX 26-SEP-2003; 2003US-0506479P.  
PR 20-JAN-2004; 2004US-0537729P.  
XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.  
XX Lahm A, Nardella C, Pallaoro M, Steinkuhler C;  
XX WPI; 2005-273382/28.  
DR N-PSDB; ADZ19001.  
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
PT nucleotide sequence that encodes mammalian heparanase protein and has two  
PT consensus cleavage sites located between specific nucleotide encoding  
PT residues.  
XX Example 2; SEQ ID NO 21; 65pp; English.  
XX



CC The invention relates to a synthetic nucleic acid molecule that encodes  
CC mammalian heparanase protein, where the nucleic acid comprises two  
CC consensus cleavage sites recognized by endoproteinase. The sequences are  
CC useful for expressing mammalian heparanase in non-mammalian cells and in  
CC inhibitor screening assays for the development of therapeutics or  
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
CC and/or inflammation. This sequence represents a hepsG3 construct protein  
CC used in the scope of the invention.

XX SQ Sequence 501 AA;

Query Match 100.0%; Score 86; DB 9; Length 501;  
Best Local Similarity 100.0%; Pred. No. 8e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14  
Db 252 TWHYYLNGRTATR 265  
|||||

RESULT 19  
ADZ19005  
ID ADZ19005 standard; protein; 507 AA.

AC ADZ19005;

XX 16-JUN-2005 (first entry)

XX HepG6 construct protein.

XX Enzyme engineering; heparanase; hepsG6; metastasis; autoimmune disease;  
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
KW immunosuppressive; enzyme.

XX Synthetic.

XX WO2005030962-A1.

XX 07-APR-2005.

XX 17-SEP-2004; 2004WO-EP010517.

XX 26-SEP-2003; 2003US-0506479P.

XX 20-JAN-2004; 2004US-0537729P.

XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.

XX Lahm A, Nardella C, Pallaoro M, Steinkuhler C;

XX WPI; 2005-273382/28.

XX N-PSDB; ADZ19003.

XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
PT nucleotide sequence that encodes mammalian heparanase protein and has two  
PT consensus cleavage sites located between specific nucleotide encoding  
PT residues.

XX Example 2; SEQ ID NO 26; 65pp; English.

XX The invention relates to a synthetic nucleic acid molecule that encodes  
CC mammalian heparanase protein, where the nucleic acid comprises two  
CC consensus cleavage sites recognized by endoproteinase. The sequences are  
CC useful for expressing mammalian heparanase in non-mammalian cells and in  
CC inhibitor screening assays for the development of therapeutics or  
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
CC and/or inflammation. This sequence represents a hepsG6 construct protein  
CC used in the scope of the invention.

XX SQ Sequence 507 AA;

Query Match 100.0%; Score 86; DB 9; Length 507;  
Best Local Similarity 100.0%; Pred. No. 8.2e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14  
Db 258 TWHYYLNGRTATR 271  
|||||

RESULT 20  
ADY27058

ID ADY27058 standard; protein; 508 AA.

XX AC ADY27058;

XX 05-MAY-2005 (first entry)

XX Human inactive heparanase protein.

XX Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
KW neurological disease; viral infection; infection; cytostatic;  
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
KW protease; enzyme; enzyme purification.

XX Homo sapiens.

XX WO2005016227-A2.

XX 24-FEB-2005.

XX 12-AUG-2004; 2004WO-IL000744.

XX 14-AUG-2003; 2003US-0494800P.

XX 12-JAN-2004; 2004US-0535492P.

XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX Van-Gelder JM, Miron D;

XX WPI; 2005-182203/19.

XX Regulating heparanase activity, useful for treating heparanase-associated  
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
PT neurological diseases or viral diseases) comprises modulating heparanase  
PT activation.

XX Claim 257; SEQ ID NO 34; 211pp; English.

XX The invention relates to a method of regulating heparanase activity in a  
CC tissue or regulating a biological process depending at least in part on  
CC heparanase activity comprising modulating heparanase activation. The  
CC invention also relates to methods of treating a heparanase- or heparin  
CC binding protein-associated disease or disorder in a subject, a  
CC pharmaceutical composition for use in the treatment of a heparanase-  
CC associated disease or disorder comprising a therapeutic amount of an  
CC agent capable of modulating heparanase activation and a pharmaceutical  
CC carrier or diluent, a method of identifying a protease activator of  
CC heparanase, a protease substrate mimetic comprising a peptide  
CC representing a subset or all substrate residues or cleavage sites of  
CC human heparanase or an equivalent non-human heparanase, a method of  
CC producing active heparanase and a method of modulating an adhesion  
CC activity of heparanase. The composition and methods are useful for  
CC modulating heparanase activation and for treating heparanase-associated  
CC diseases or disorders such as cancer, inflammation, cardiovascular  
CC diseases, neurological diseases or viral infections. This sequence  
CC represents a human inactive heparanase protein used in the scope of the  
CC invention.

XX SQ Sequence 508 AA;

Query Match 100.0%; Score 86; DB 9; Length 508;  
Best Local Similarity 100.0%; Pred. No. 8.2e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14  
|||||

Db 259 TWHYYLNGRTATR 272

RESULT 21  
ADZ19006  
ID ADZ19006 standard; protein; 526 AA.  
XX  
AC ADZ19006;  
XX  
DT 16-JUN-2005 (first entry)  
XX  
DE HephYaluro construct protein.  
XX  
KW Enzyme engineering; heparanase; hephYaluro; metastasis;  
KW autoimmune disease; inflammation; neoplasm; immune disorder;  
KW antiinflammatory; cytostatic; immunosuppressive; enzyme.  
XX  
OS Synthetic.  
XX  
PN WO2005030962-A1.  
XX  
PD 07-APR-2005.  
XX  
PF 17-SEP-2004; 2004WO-EP010517.  
XX  
PR 26-SEP-2003; 2003US-0506479P.  
PR 20-JAN-2004; 2004US-0537729P.  
XX  
PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.  
XX  
PI Lahm A, Nardella C, Pallaoro M, Steinkuhler C;  
XX  
DR WPI: 2005-273382/28.  
DR N-PSDB; ADZ19007.  
XX  
PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
PT nucleotide sequence that encodes mammalian heparanase protein and has two  
PT consensus cleavage sites located between specific nucleotide encoding  
PT residues.  
XX  
PS Example 2; SEQ ID NO 27; 65pp; English.  
XX  
CC The invention relates to a synthetic nucleic acid molecule that encodes  
CC mammalian heparanase protein, where the nucleic acid comprises two  
CC consensus cleavage sites recognized by endoproteinase. The sequences are  
CC useful for expressing mammalian heparanase in non-mammalian cells and in  
CC inhibitor screening assays for the development of therapeutics or  
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
CC and/or inflammation. This sequence represents a hephYaluro construct  
CC protein used in the scope of the invention.  
XX  
SQ Sequence 526 AA;  
Query Match 100.0%; Score 86; DB 9; Length 526;  
Best Local Similarity 100.0%; Pred. No. 8.5e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TWHYYLNGRTATR 14  
|||||  
Db 277 TWHYYLNGRTATR 290  
RESULT 22  
ABB07815  
ID ABB07815 standard; protein; 527 AA.  
XX  
AC ABB07815;  
XX  
DT 03-JUL-2002 (first entry)  
XX  
DE Chicken signal peptide/human heparanase chimeric protein sequence.  
XX  
KW Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;  
anti-protozoan; neuroprotective; heparin; chicken; human; chimeric.  
Synthetic.  
OS Gallus gallus.  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..19  
FT Protein /note= "chicken heparanase signal peptide"  
FT /note= "human heparanase mature protein"  
XX  
PN US2002034810-A1.  
XX  
PD 21-MAR-2002.  
XX  
PF 16-AUG-2001; 2001US-00930218.  
XX  
PR 20-SEP-2000; 2000US-00666390.  
XX  
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
XX  
PI Goldshmidt O, Pecker I, Vlodaysky I, Michal I, Zcharia E;  
XX  
DR WPI: 2002-338926/37.  
DR N-PSDB; ABL40753.  
XX  
PT Nucleic acid encoding avian and reptile heparanase polypeptide is useful  
PT to treat various heparin-related disorders and the signal peptide is  
PT useful in production of membrane-targeted or secreted recombinant  
PT proteins.  
XX  
PS Disclosure; Page 26-28; 39pp; English.  
XX  
CC The invention relates to an isolated avian and reptile nucleic acid,  
CC encoding a polypeptide with heparanase catalytic activity. The signal  
CC peptide of the nucleic acid can be used to express membrane-associated or  
CC secreted proteins in heterologous expression systems. The encoded  
CC polypeptides can be used to prevent tumour angiogenesis, metastasis and  
CC invasion, and to intervene with pathologies associated with impaired  
CC heparin-binding growth factors, cellular responses to heparin-binding  
CC growth factors and cytokines, cell interaction with plasma lipoproteins,  
CC cellular susceptibility to viral, protozoa and bacterial infections or  
CC disintegration of neurodegenerative plaques. The present sequence  
CC represents a chicken signal peptide/human heparanase chimeric protein  
CC sequence  
XX  
SQ Sequence 527 AA;  
Query Match 100.0%; Score 86; DB 5; Length 527;  
Best Local Similarity 100.0%; Pred. No. 8.5e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TWHYYLNGRTATR 14  
|||||  
Db 278 TWHYYLNGRTATR 291  
RESULT 23  
ABW02018  
ID ABW02018 standard; protein; 527 AA.  
XX  
AC ABW02018;  
XX  
DT 12-FEB-2004 (first entry)  
XX  
DE Chimeric human-chicken heparanase protein.  
XX  
KW Chicken; heparanase; tumour cell metastasis; inflammation; autoimmunity;  
KW wound healing; angiogenesis; restenosis; Genstamm-Straussler Syndrome;  
KW neurodegenerative disease; atherosclerosis; Creutzfeldt-Jakob disease;  
KW infection; Scrapie; Alzheimer's disease; protein therapy; cytostatic;  
KW immunosuppressive; vulnerary; bactericide; anti-angiogenic; virucide;

KW antisclerotic; neuroprotective; protozoacide; chimeric; fusion protein;  
 KW enzyme; human.  
 XX  
 OS Chimeric - Gallus gallus.  
 OS Chimeric - Homo sapiens.  
 XX  
 PN US2003180788-A1.  
 XX  
 XX 25-SEP-2003.  
 PD  
 XX 08-MAY-2003; 2003US-00431438.  
 PF  
 XX 20-SEP-2000; 2000US-00666390.  
 PR  
 PR 16-AUG-2001; 2001US-00930218.  
 XX  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX  
 XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;  
 PI  
 XX WPI; 2003-843931/78.  
 DR N-PSDB; AAD63532.  
 XX  
 PT Recombinant jungle red fowl (Gallus gallus) heparanase protein, useful  
 PT for treating cancers, microbial infections and aiding wound healing.  
 XX  
 PS Example; Page 26-28; 0pp; English.  
 XX  
 CC The present invention relates to novel jungle red fowl heparanase protein  
 CC and polynucleotides encoding such proteins. Heparanase sequences can be  
 CC used to develop treatments for various diseases, to develop diagnostic  
 CC assays for these diseases and to provide new tools for basic and directed  
 CC research especially in the fields of medicine and biology. They can be  
 CC used to develop new drugs to inhibit tumour cell metastasis, inflammation  
 CC and autoimmunity. Recombinant heparanase offers a potential treatment for  
 CC wound healing, angiogenesis, restenosis, atherosclerosis, inflammation,  
 CC neurodegenerative diseases (e.g. Genetmann-Straussler Syndrome, Scrape,  
 CC Creutzfeldt-Jakob disease and Alzheimer's disease) and certain viral and  
 CC some bacterial and protozoa infections. Recombinant heparanase can also  
 CC be used to neutralise plasma heparin, as a potential replacement of  
 CC protamine. Sequences of the invention are used in protein therapy. The  
 CC present sequence is chimeric human-chicken heparanase protein  
 XX  
 SQ Sequence 527 AA;  
 Query Match 100.0%; Score 86; DB 7; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 TWHYYLNGRTATR 14  
 Db 278 TWHYYLNGRTATR 291  
 RESULT 24  
 ADO63825  
 ID ADO63825 standard; protein; 527 AA.  
 XX  
 AC ADO63825;  
 XX  
 XX 26-AUG-2004 (first entry)  
 DT  
 XX Chimeric heparanase mutant E225A, SEQ ID:10.  
 DE  
 XX Human; chicken; heparanase; heparanase-derived protein;  
 KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;  
 KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;  
 KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnuary; mutant; mutein.  
 XX  
 OS Homo sapiens.

OS Gallus gallus.  
 OS Synthetic.  
 OS Chimeric.  
 XX  
 FH Key  
 FT Peptide  
 FT  
 FT Region  
 FT  
 FT Misc-difference 209  
 FT /note= "Ala replaces wild-type Glu (active site proton  
 FT donor). Corresponds to residue 225 of human heparanase  
 FT mutant E225A (SEQ ID NO:7) "  
 FT 327  
 FT /note= "Active site nucleophile"  
 XX  
 PN WO2004048558-A2.  
 XX  
 XX 10-JUN-2004.  
 PD  
 XX 24-NOV-2003; 2003WO-IL000989.  
 PF  
 PR 24-NOV-2002; 2002IL-00153059.  
 XX  
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA  
 XX Vlodavsky I, Zcharia E, Goldshmidt O, Ilan N;  
 PI  
 XX WPI; 2004-450373/42.  
 DR N-PSDB; ADO63819.  
 XX  
 CC New nucleic acid construct comprising heparanase-derived polypeptide,  
 CC useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 CC thrombasthenia, or Bernard-Soulier syndrome.  
 PT  
 PT Claim 10; SEQ ID NO 10; 128pp; English.  
 PS  
 XX The invention relates to nucleic acid constructs comprising a nucleic  
 XX acid encoding a heparanase-derived protein which lacks heparanase  
 CC endoglycosidase catalytic activity but which retains its cell-cell and  
 CC cell-matrix adhesion properties. The constructs of the invention  
 CC optionally further comprise operably linked regulatory elements. The  
 CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the  
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may (ADO63825-  
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
 CC host cells are useful in preparing a tissue sealant composition for  
 CC sealing injuries, reducing the loss of blood, accelerating the healing  
 CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialisation of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be use in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte  
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents a chimeric  
 CC protein comprising the signal peptide of chicken heparanase and residues  
 CC 35-543 of the human heparanase mutant E225A.  
 XX  
 SQ Sequence 527 AA;  
 Query Match 100.0%; Score 86; DB 8; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 TWHYYLNGRTATR 14

Db	278 TWHHYLLNGRTATR 291	
RESULT 25		
AD063827		
ID	AD063827 standard; protein; 527 AA.	
XX		
AC	AD063827;	
XX		
DT	26-AUG-2004 (first entry)	
XX		
DE	Chimeric heparanase mutant E225A/E343A, SEQ ID:12.	
XX		
KW	Human; chicken; heparanase; heparanase-derived protein;	
KW	heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;	
KW	matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;	
KW	blood vessel; vascular graft; platelet adhesion; platelet aggregation;	
KW	adhesion disorder; LAD; leukocyte adhesion deficiency; drug screening;	
KW	Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;	
KW	vulnerary; mutant; mutein.	
XX		
OS	Homo sapiens.	
OS	Gallus gallus.	
OS	Synthetic.	
OS	Chimeric.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1. .18
FT	Region	/note= "Chicken heparanase signal peptide"
FT		19. .527
FT		/note= "Corresponds to residues 35-543 of human
FT		heparanase mutant E225A/E343A (SEQ ID NO:9)"
FT	Misc-difference 209	
FT		/note= "Ala replaces wild-type Glu (active site proton
FT		donor). Corresponds to residue 225 of human heparanase
FT		mutant E225A/E343A (SEQ ID NO:9) "
FT	Misc-difference 327	
FT		/note= "Ala replaces wild-type Glu (active site
FT		nucleophile). Corresponds to residue 343 of human
FT		heparanase mutant E225A/E343A (SEQ ID NO:9) "
XX		
PN	WO2004048558-A2.	
XX		
PD	10-JUN-2004.	
XX		
PF	24-NOV-2003; 2003WO-IL000989.	
XX		
PR	24-NOV-2002; 2002IL-00153059.	
XX		
PA	(HADA-) HADASIT MEDICAL RES SERVICES & DEV.	
XX		
PI	Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;	
XX		
DR	WPI; 2004-450373/42.	
DR	N-PSDB; AD063821.	
XX		
PT	New nucleic acid construct comprising heparanase-derived polypeptide,	
PT	useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's	
PT	thrombasthenia, or Bernard-Soulier syndrome.	
XX		
PS	Claim 10; SEQ ID NO 12; 128pp; English.	
XX		
CC	The invention relates to nucleic acid constructs comprising a nucleic	
CC	acid encoding a heparanase-derived protein which lacks heparanase	
CC	endoglycosidase catalytic activity but which retains its cell-cell and	
CC	cell-matrix adhesion properties. The constructs of the invention	
CC	optionally further comprise operably linked regulatory elements. The	
CC	invention also relates to the heparanase-derived proteins and host cells	
CC	comprising the nucleic acid constructs of the invention. The heparanase-	
CC	derived proteins are especially mutants of human heparanase in which the	
CC	active site proton donor Glu225 and/or the active site nucleophile Glu343	
CC	are replaced with Ala (AD063822-AD063824), and the proteins may	
CC		
CC	optionally further comprise an avian heparanase signal peptide (AD063825-	
CC	AD063827). The heparanase-derived protein, nucleic acid construct and	
CC	host cells are useful in preparing a tissue sealant composition for	
CC	sealing injuries, reducing the loss of blood, accelerating the healing	
CC	and homeostasis of an injury, accelerating blood vessel endothelium	
CC	formation or the endothelialisation of vascular grafts, accelerating the	
CC	adhesive activity of mammalian cells, and accelerating the adhesion and	
CC	aggregation of platelets. They may also be used in the treatment of	
CC	disorders associated with adhesion deficiency such as LAD (leukocyte	
CC	adhesion deficiency), Glanzmann's thrombasthenia (defective platelet	
CC	function), or Bernard-Soulier syndrome (deficient platelet adhesion). The	
CC	cells of the invention may additionally be used to screen for modulators of	
CC	cell-cell and cell-matrix adhesion, and to prepare an implantable	
CC	synthetic vascular graft comprising a tube made of a biocompatible	
CC	material lined with the cells. The present sequence represents a chimeric	
CC	protein comprising the signal peptide of chicken heparanase and residues	
CC	35-543 of the human heparanase double mutant E225A/E343A.	
XX		
SEQ	Sequence 527 AA;	
	Query Match 100.0%; Score 86; DB 8; Length 527;	
	Best Local Similarity 100.0%; Pred. No. 8.5e-06;	
	Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 TWHHYLLNGRTATR 14	
DB	278 TWHHYLLNGRTATR 291	
RESULT 26		
AD063826		
ID	AD063826 standard; protein; 527 AA.	
XX		
AC	AD063826;	
XX		
DT	26-AUG-2004 (first entry)	
XX		
DE	Chimeric heparanase mutant E343A, SEQ ID:11.	
XX		
KW	Human; chicken; heparanase; heparanase-derived protein;	
KW	heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;	
KW	matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;	
KW	blood vessel; vascular graft; platelet adhesion; platelet aggregation;	
KW	adhesion disorder; LAD; leukocyte adhesion deficiency;	
KW	Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;	
KW	vulnerary; mutant; mutein.	
XX		
OS	Homo sapiens.	
OS	Gallus gallus.	
OS	Synthetic.	
OS	Chimeric.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1. .18
FT	Region	/note= "Chicken heparanase signal peptide"
FT		19. .527
FT		/note= "Corresponds to residues 35-543 of human
FT		heparanase mutant E343A (SEQ ID NO:8)"
FT	Active-site 209	
FT		/note= "Active site proton donor"
FT	Misc-difference 327	
FT		/note= "Ala replaces wild-type Glu (active site
FT		nucleophile). Corresponds to residue 343 of human
FT		heparanase mutant E343A (SEQ ID NO:8) "
XX		
XX	WO2004048558-A2.	
XX		
XX	10-JUN-2004.	
XX		
XX	24-NOV-2003; 2003WO-IL000989.	
XX		
XX	24-NOV-2002; 2002IL-00153059.	
XX		
XX	(HADA-) HADASIT MEDICAL RES SERVICES & DEV.	
XX		
XX	Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;	
XX		
XX	WPI; 2004-450373/42.	
XX	N-PSDB; AD063821.	
XX		
XX	New nucleic acid construct comprising heparanase-derived polypeptide,	
XX	useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's	
XX	thrombasthenia, or Bernard-Soulier syndrome.	
XX		
XX	Claim 10; SEQ ID NO 12; 128pp; English.	
XX		
XX	The invention relates to nucleic acid constructs comprising a nucleic	
XX	acid encoding a heparanase-derived protein which lacks heparanase	
XX	endoglycosidase catalytic activity but which retains its cell-cell and	
XX	cell-matrix adhesion properties. The constructs of the invention	
XX	optionally further comprise operably linked regulatory elements. The	
XX	invention also relates to the heparanase-derived proteins and host cells	
XX	comprising the nucleic acid constructs of the invention. The heparanase-	
XX	derived proteins are especially mutants of human heparanase in which the	
XX	active site proton donor Glu225 and/or the active site nucleophile Glu343	
XX	are replaced with Ala (AD063822-AD063824), and the proteins may	
XX		

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
 PI  
 XX  
 DR WPI; 2004-450373/42.  
 DR N-PSDB; ADO63820.  
 XX  
 PT New nucleic acid construct comprising heparanase-derived polypeptide,  
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 PT thrombasthenia, or Bernard-Soulier syndrome.  
 XX  
 PS Claim 10; SEQ ID NO 11; 128pp; English.  
 XX  
 CC The invention relates to nucleic acid constructs comprising a nucleic  
 CC acid encoding a heparanase-derived protein which lacks heparanase  
 CC endoglycosidase catalytic activity but which retains its cell-cell and  
 CC cell-matrix adhesion properties. The constructs of the invention  
 CC optionally further comprise operably linked regulatory elements. The  
 CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the  
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
 CC host cells are useful in preparing a tissue sealant composition for  
 CC sealing injuries, reducing the loss of blood, accelerating the healing  
 CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialisation of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be used in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte  
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents a chimeric  
 CC protein comprising the signal peptide of chicken heparanase and residues  
 CC 35-543 of the human heparanase mutant E343A.  
 XX  
 SQ Sequence 527 AA;  
 Query Match 100.0%; Score 86; DB 8; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 TWHYYLNGRTATR 14  
 |||||  
 Db 278 TWHYYLNGRTATR 291  
 RESULT 27  
 AD219004  
 ID AD219004 standard; protein; 527 AA.  
 AC  
 AC AD219004;  
 XX  
 DT 16-JUN-2005 (first entry)  
 DE HepGS4 construct protein.  
 XX  
 KW Enzyme engineering; heparanase; hepGS4; metastasis; autoimmune disease;  
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
 KW immunosuppressive; enzyme.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2005030962-A1.  
 PD 07-APR-2005.  
 XX  
 PP 17-SEP-2004; 2004WO-EP010517.

XX 26-SEP-2003; 2003US-0506479P.  
 PR 20-JAN-2004; 2004US-0537729P.  
 XX  
 PA (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.  
 XX  
 PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;  
 XX  
 DR WPI; 2005-273382/28.  
 DR N-PSDB; ADZ19002.  
 XX  
 PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.  
 PS Example 2; SEQ ID NO 25; 65pp; English.  
 XX  
 CC The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoproteinase. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in  
 CC inhibitor screening assays for the development of therapeutics or  
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a hepGS4 construct protein  
 CC used in the scope of the invention.  
 XX  
 SQ Sequence 527 AA;  
 Query Match 100.0%; Score 86; DB 9; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-06; Indels 0; Gaps 0;  
 Matches 14; Conservative 0; Mismatches 0;  
 OY 1 TWHYYLNGRTATR 14  
 |||||  
 Db 278 TWHYYLNGRTATR 291  
 RESULT 28  
 AAY17083  
 ID AAY17083 standard; protein; 532 AA.  
 XX  
 AC AAY17083;  
 XX  
 DT 21-JUL-1999 (first entry)  
 XX  
 DE Seq ID No: 15 of WO9921975.  
 XX  
 KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;  
 KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;  
 KW arteriosclerosis; atherosclerosis; inflammation; tissue development;  
 KW human; HSPG.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9921975-A1.  
 XX  
 PD 06-MAY-1999.  
 XX  
 PP 28-OCT-1998; 98WO-AU000898.  
 XX  
 PR 28-OCT-1997; 97AU-00000062.  
 PR 09-DEC-1997; 97AU-00000812.  
 XX  
 XX (AUSU) UNIV AUSTRALIAN NAT.  
 FA  
 XX Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;  
 PI  
 XX WPI; 1999-312956/26.  
 DR N-PSDB; AAX37260.  
 XX  
 PT Polynucleotides encoding mammalian endoglucuronidases, especially  
 PT heparanases, useful to promote wound healing.

XX Claim 6; Page 76-79; 112pp; English.

PS The invention relates to nucleic acid sequences that encode heparanase

CC enzymes having endoglucuronidase activity. Recombinant heparanases are

CC capable of removing the HS side chain from heparan sulfate proteoglycan

CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to

CC inhibit heparanase, this is useful for treatment of a physiological or

CC medical condition associated with elevated heparanase activity, such as

CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,

CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and

CC rat heparanases can be used to enhance wound healing, especially

CC associated with tissue development and repair. The conditions mentioned

CC above can be diagnosed using specific antibodies, and also using primers

CC and probes specific for the heparanase polynucleotides. Other uses of the

CC heparanases include sequencing sulfated molecules such as HSPG

XX

SQ Sequence 532 AA;

Query Match 100.0%; Score 86; DB 2; Length 532;

Best Local Similarity 100.0%; Pred. No. 8.6e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14

DB 294 TWHYYLNGRTATR 307

|||||

RESULT 29

AA02345

ID AAY02345 standard; protein; 543 AA.

XX

AC AAY02345;

XX

DT 09-JUL-1999 (first entry)

XX

DE A human heparanase protein.

XX

XX Heparanase; hp; modulator; heparin-binding growth factor;

XX cellular response; cytokine; cell interaction; plasma lipoprotein;

KW cellular susceptibility; infection; disintegration;

KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;

KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;

KW plasma heparin; micrometastasis; autoimmune lesion; renal failure.

XX

OS Homo sapiens.

XX

PN WO9911798-A1.

XX

PD 11-MAR-1999.

XX

PF 31-AUG-1998; 98WO-US017954.

XX

PR 02-SEP-1997; 97US-00922170.

PR

PR 02-JUL-1998; 98US-00109386.

XX

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

PA (FRIE/) FRIEDMAN M M.

XX

XX Pecker I, Vlodavsky I, Feinstein E;

XX

XX WPI; 1999-302255/25.

DR

DR N-PSDB; AAX35648.

XX

XX New human polynucleotide useful for treating angiogenesis, restenosis,

PT and inflammation.

XX

PS Claim 6; Fig 1; 63pp; English.

XX

XX The specification describes a polypeptide having heparanase (hp)

CC activity. The recombinant protein is used as a modulator of heparin-

CC binding growth factors, cellular responses to heparin-binding growth

CC factors and cytokines, cell interaction with plasma lipoproteins,

CC cellular susceptibility to viral, protozoal and bacterial infections or

CC disinfection of neurodegenerative plaques. Heparanase may be useful for

CC conditions such as wound healing, angiogenesis, restenosis,

CC atherosclerosis, inflammation, neurodegenerative diseases, and viral

CC infections. Mammalian heparanase can be used to neutralize plasma

CC heparin, and anti-heparanase antibodies may be applied for

CC immunodetection and diagnosis of micrometastases, autoimmune lesions, and

CC renal failure in biopsy specimens, plasma samples, and body fluids. The

CC present sequence represents human heparanase

XX

SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 2; Length 543;

Best Local Similarity 100.0%; Pred. No. 8.8e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14

DB 294 TWHYYLNGRTATR 307

|||||

RESULT 30

AA017082

ID AAY17082 standard; protein; 543 AA.

XX

AC AAY17082;

XX

DT 21-JUL-1999 (first entry)

XX

DE Human heparanase enzyme.

XX

XX Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;

KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;

KW arteriosclerosis; atherosclerosis; inflammation; tissue development;

KW human; HSPG.

XX

OS Homo sapiens.

XX

PN WO9921975-A1.

XX

PD 06-MAY-1999.

XX

PF 28-OCT-1998; 98WO-AU000898.

XX

PR 28-OCT-1997; 97AU-00000062.

PR

PR 09-DEC-1997; 97AU-00000812.

XX

XX (AUSU) UNIV AUSTRALIAN NAT.

XX

XX Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;

PI

XX WPI; 1999-312956/26.

DR

DR N-PSDB; AAX37259.

XX

XX Polynucleotides encoding mammalian endoglucuronidases, especially

PT heparanases, useful to promote wound healing.

XX

PS Claim 6; Page 69-73; 112pp; English.

XX

XX The invention relates to nucleic acid sequences that encode heparanase

CC enzymes having endoglucuronidase activity. Recombinant heparanases are

CC capable of removing the HS side chain from heparan sulfate proteoglycan

CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to

CC inhibit heparanase, this is useful for treatment of a physiological or

CC medical condition associated with elevated heparanase activity, such as

CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,

CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and

CC rat heparanases can be used to enhance wound healing, especially

CC associated with tissue development and repair. The conditions mentioned

CC above can be diagnosed using specific antibodies, and also using primers

CC and probes specific for the heparanase polynucleotides. Other uses of the

CC heparanases include sequencing sulfated molecules such as HSPG. The

CC present sequence represents a human heparanase  
 XX  
 SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 2; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14  
 |||||  
 Db 294 TWHYYLNGRTATR 307

RESULT 31  
 AAY57590  
 ID AAY57590 standard; protein; 543 AA.  
 XX  
 AC AAY57590;  
 XX  
 DT 02-MAR-2000 (first entry)  
 XX  
 DE Human heparanase.  
 XX  
 KW Human; heparanase; hpa; genetic modification; expression; anticancer;  
 KW angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumour;  
 KW anti-atherosclerotic; anti-inflammatory; antineurodegeneration;  
 KW heparan sulphate; heparin-binding growth factor; tumour angiogenesis;  
 KW metastasis; wound healing; restenosis; atherosclerosis; inflammation;  
 KW neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;  
 KW micrometastasis; autoimmune lesion; kidney failure.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9957244-A1.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 29-APR-1999; 99WO-US009256.  
 XX  
 PR 01-MAY-1998; 98US-00071618.  
 PR 02-MAR-1999; 99US-00260038.  
 XX  
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 (FRIE/) FRIEDMAN M M.  
 XX  
 PI Ben-Artzi H, Ayal-Herskovitz M, Yacoby-Zeevi O, Pecker I;  
 PI Peleg Y, Shlomi Y;  
 XX  
 WPI; 2000-062144/05.  
 DR N-PSDB; AA239195.  
 XX  
 PT Engineered cells that express recombinant heparanase, useful  
 PT therapeutically, e.g. for treating angiogenesis and to screen for  
 PT specific inhibitors, potential anticancer agents.  
 XX  
 PS Claim 3; Page 107-109; 118pp; English.  
 XX  
 CC The present invention describes genetically modified cells (A) containing  
 CC a polynucleotide (I) that encodes a polypeptide with heparanase activity,  
 CC and express recombinant heparanase (II). Heparanase cleaves heparan  
 CC sulphate (HS) at specific intrachain sites, resulting in release of  
 CC heparin-binding growth factors, enzymes and proteins that are sequestered  
 CC by HS in basement membranes, extracellular matrix or cell surfaces. It  
 CC may also be implicated in tumour angiogenesis and metastases. (II) is  
 CC potentially useful in wound healing and for treating angiogenesis,  
 CC restenosis, atherosclerosis, inflammation, neurodegeneration, viral  
 CC infection and cystic fibrosis. It can also be used to neutralise heparin  
 CC (an alternative to protamine) and to screen for specific inhibitors  
 CC (potentially useful for treating cancer and metastases). Antibodies  
 CC raised against (II) are used for immunodetection and diagnosis of  
 CC micrometastases, autoimmune lesions and kidney failure. (A) provide (II)  
 CC in large quantities, in a form that is homogeneously processed and  
 CC activated/neutralised by a dedicated protease. The present sequence

CC represents human heparanase  
 XX  
 SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 3; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14  
 |||||  
 Db 294 TWHYYLNGRTATR 307

RESULT 32  
 AAB08849  
 ID AAB08849 standard; protein; 543 AA.  
 XX  
 AC AAB08849;  
 XX  
 DT 15-JAN-2001 (first entry)  
 XX  
 DE Amino acid sequence of a human heparanase polypeptide.  
 XX  
 KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;  
 KW heparin-binding growth factor; cytokine; neurodegenerative plaque;  
 KW wound healing; infection; burn; angiogenesis; restenosis;  
 KW atherosclerosis; inflammation; neurodegenerative disease;  
 KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200052178-A1.  
 XX  
 PD 08-SEP-2000.  
 XX  
 PF 14-FEB-2000; 2000WO-US003542.  
 XX  
 PR 01-MAR-1999; 99US-00258892.  
 XX  
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA (FRIE/) FRIEDMAN M M.  
 XX  
 PI Pecker I, Vlodavsky I, Feinstein E;  
 XX  
 DR WPI; 2000-579289/54.  
 DR N-PSDB; AAA75051.  
 XX  
 PT New polynucleotides encoding a polypeptide having heparanase activity,  
 PT useful in wound healing and in gene therapy, particularly in treating  
 PT tumor, inflammation, autoimmunity, neurodegenerative diseases.  
 XX  
 PS Claim 22; Fig 1; 152pp; English.  
 XX  
 CC The present sequence represents a human protein with heparanase catalytic  
 CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,  
 CC particularly in treating tumour, inflammation or autoimmunity.  
 CC Particularly, the polynucleotide is useful in modulating the  
 CC bioavailability of heparin-binding growth factors, cellular responses to  
 CC heparin-binding growth factors (e.g. bFGF) and cytokines (e.g.  
 CC interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular  
 CC susceptibility to certain viral and some bacterial and protozoa  
 CC infections, or disintegration of neurodegenerative plaques. The  
 CC polynucleotide is also useful in wound healing (e.g. thermal, chemical or  
 CC radiation burns), and in the treatment of angiogenesis, restenosis,  
 CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-  
 CC Straussler Syndrome or Creutzfeldt-Jakob disease), and some viral,  
 CC bacterial or protozoa infections  
 XX  
 SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 3; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-06;



```
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
Db 294 TWHYYLNGRTATR 307

RESULT 33
AAY52990
ID AAY52990 standard; protein; 543 AA.
XX
AC AAY52990;
XX
DT 21-FEB-2000 (first entry)
XX
DE Human heparanase protein sequence.
XX
KW Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;
KW anti-diabetic; immunomodulatory; anti-inflammatory; nephrotropic;
KW metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;
KW mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;
KW inflammation; haemorrhagic nephritis; nephrotic syndrome;
KW autoimmune disease; anticancer; kidney disease.
XX
OS Homo sapiens.
XX
PN WO9957153-A1.
XX
PD 11-NOV-1999.
XX
PF 29-APR-1999; 99WO-US009255.
XX
PR 01-MAY-1998; 98US-00071739.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
PI Pecker I, Vlodavsky I, Friedman Y, Perets T;
XX
DR WPI; 2000-052944/04.
DR N-PSDB; AA233290.
XX
PT Heparanase-specific molecular probes useful for diagnosis and treatment,
PT e.g. of tumors, and for targeted drug delivery.
XX
PS Example; Page 81-82; 90pp; English.
XX
CC The present invention describes heparanase-specific molecular probes,
CC useful for methods of detecting heparanase in situ. The probes and anti-
CC heparanase antibodies are used to detect or quantify the expression of
CC heparanase, for diagnosis and monitoring of diseases (especially
CC metastasis), for treatment of heparanase-associated diseases (e.g.
CC tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,
CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its
CC metastases) derived from liver, prostate, bladder, breast, ovary, cervix,
CC colon, skin, intestine, stomach, uterus and pancreas, kidney disease,
CC diabetes and inflammation, haemorrhagic nephritis, nephrotic syndrome,
CC sepsis and inflammatory or autoimmune disease), for targeted drug
CC delivery (e.g. of anticancer agents) and as research reagents. The
CC present sequence represents human heparanase, which is used in the
CC exemplification of the present invention
XX
SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 3; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
Db 294 TWHYYLNGRTATR 307

RESULT 34
AAY97635
ID AAY97635 standard; protein; 543 AA.
XX
AC AAY97635;
XX
DT 20-APR-2001 (first entry)
XX
DE Human heparanase protein sequence.
XX
KW Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW gene therapy; human.
XX
OS Homo sapiens.
XX
PN WO200100643-A2.
XX
PD 04-JAN-2001.
XX
PF 19-JUN-2000; 2000WO-IL000358.
XX
PR 25-JUN-1999; 99US-0140801P.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA Pecker I, Michal I, Itzhaki H;
XX
DR WPI; 2001-137930/14.
XX
PT New polynucleotides and polypeptides that are distantly homologous to
PT heparanase, useful in wound healing, as well as in gene therapy protocols
PT for angiogenesis, restenosis, atherosclerosis, or inflammation.
XX
PS Disclosure; Page 64-65; 67pp; English.
XX
CC This sequence represents a heparanase of the invention. The heparanase
CC DNA and protein sequences are useful in wound healing, angiogenesis,
CC restenosis, atherosclerosis, inflammation, pulmonary diseases,
CC neurodegenerative diseases (such as Scrape, Alzheimer's disease, and
CC Creutzfeldt-Jakob disease) or viral infections. The heparanase coding
CC sequence is particularly useful in gene therapy
XX
SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 4; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
Db 294 TWHYYLNGRTATR 307

RESULT 35
AAB86206
ID AAB86206 standard; protein; 543 AA.
XX
AC AAB86206;
XX
DT 24-AUG-2001 (first entry)
XX
DE Human heparanase inhibitor protein.
XX
KW Heparanase; inhibitor; cardiac insufficiency; cardiant; nephrotropic;
KW hepatotropic; veterinary medicine; congestive heart failure; dyspnoea;
KW primary cardiomyopathy; peripheral edema; pulmonary congestion;
KW hepatic congestion; hydrothorax; ascite; nocturia; human.
XX
OS Homo sapiens.
XX
```

PN DE19955803-A1.  
 XX  
 PD 23-MAY-2001.  
 XX  
 PF 19-NOV-1999; 99DE-01055803.  
 XX  
 PR 19-NOV-1999; 99DE-01055803.  
 XX  
 PA (KNOL ) KNOLL AG.  
 XX  
 PI Herr D, Hahn A, Laux V;  
 XX  
 XX WPI; 2001-368371/39.  
 DR N-PSDB; AAF20940.  
 DR  
 XX  
 PT Treatment or prevention of cardiac insufficiency and related conditions,  
 PT e.g. pulmonary congestion and dyspnoea, comprises administration of  
 PT heparanase inhibitor.  
 XX  
 PS Disclosure; Page 11-13; 16pp; German.  
 XX  
 CC This invention describes a novel heparanase inhibitor which can be used  
 CC for the treatment or prevention of cardiac insufficiency and associated  
 CC indications, symptoms and/or malfunctions. The heparanase inhibitor of  
 CC the invention has cardiac, nephrotropic and hepatotropic activity. The  
 CC products of the invention can be used in human and veterinary medicine,  
 CC for the treatment or prevention of congestive heart failure e.g. primary  
 CC cardiomyopathy. Associated conditions treated or prevented with the  
 CC inhibitor are especially peripheral odemas, pulmonary and hepatic  
 CC congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g.  
 CC nocturia can also be treated. This sequence represents the human  
 CC heparanase protein described in the method of the invention  
 XX  
 SQ Sequence 543 AA;  
 Query Match 100.0%; Score 86; DB 4; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TWHYYLNGRTATR 14  
 |||||  
 Db 294 TWHYYLNGRTATR 307  
 RESULT 36  
 AAB88361  
 ID AAB88361 standard; protein; 543 AA.  
 XX  
 AC AAB88361;  
 XX  
 DT 23-MAY-2001 (first entry)  
 XX  
 DE Human membrane or secretory protein clone PSEC0090.  
 XX  
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;  
 KW rheumatoid arthritis; diabetes.  
 XX  
 OS Homo sapiens.  
 XX  
 FN EP1067182-A2.  
 XX  
 PD 10-JAN-2001.  
 XX  
 PF 07-JUL-2000; 2000EP-00114090.  
 XX  
 PR 08-JUL-1999; 99JP-00194179.  
 PR 11-JAN-2000; 2000JP-00118775.  
 PR 02-MAY-2000; 2000JP-00183766.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
 XX

DR WPI; 2001-093989/11.  
 XX N-PSDB; AAF93788.  
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in  
 XX gene therapy or as candidate target molecules in drug development.  
 PS Claim 1; SEQ ID NO 90; 609pp + Sequence Listing; English.  
 XX  
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916  
 CC which encode human secretory or membrane proteins represented by AAB88317  
 CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and  
 CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the  
 CC invention. The invention also includes methods for the production of  
 CC antibodies directed against the proteins, and cDNA sequences, which can  
 CC be used in vaccines. The polynucleotide sequences can be used in gene  
 CC therapy. The polynucleotide sequences and the proteins they encode may be  
 CC used in the prevention, treatment and diagnosis of diseases associated  
 CC with inappropriate secretory protein/membrane protein expression. The  
 CC nucleic acids and complementary sequences may also be used as DNA probes  
 CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect  
 CC and quantitate the presence of similar nucleic acid sequences in samples.  
 CC They may also be used to study the expression and function of secretory  
 CC proteins/membrane polypeptides and their role in metabolism. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC against them and in assays to identify modulators (agonists and  
 CC antagonists) of expression and activity. The antibodies and antagonists  
 CC may also be used as therapeutic agents to down regulate expression and  
 CC activity. The antibodies may also be used as diagnostic agents for  
 CC detecting the presence of the polypeptides in samples (e.g. by enzyme  
 CC linked immunosorbant assay (ELISA). Examples of diseases which may be  
 CC treated include rheumatoid arthritis and diabetes  
 XX  
 SQ Sequence 543 AA;  
 Query Match 100.0%; Score 86; DB 4; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TWHYYLNGRTATR 14  
 |||||  
 Db 294 TWHYYLNGRTATR 307  
 RESULT 37  
 ABB07813  
 ID ABB07813 standard; protein; 543 AA.  
 XX  
 AC ABB07813;  
 XX  
 DT 03-JUL-2002 (first entry)  
 XX  
 DE Human heparanase sequence.  
 XX  
 KW Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;  
 KW anti-protozoan; neuroprotective; heparin; human.  
 XX  
 OS Homo sapiens.  
 XX  
 FN US2002034810-A1.  
 XX  
 PD 21-MAR-2002.  
 XX  
 PF 16-AUG-2001; 2001US-00930218.  
 XX  
 PR 20-SEP-2000; 2000US-00666390.  
 XX  
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;  
XX WPI; 2002-338926/37.  
XX  
XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful  
PT to treat various heparin-related disorders and the signal peptide is  
PT useful in production of membrane-targeted or secreted recombinant  
PT proteins.  
XX  
XX Disclosure; Fig 1a; 39pp; English.  
XX  
XX The invention relates to an isolated avian and reptile nucleic acid,  
CC encoding a polypeptide with heparanase catalytic activity. The signal  
CC peptide of the nucleic acid can be used to express membrane-associated or  
CC secreted proteins in heterologous expression systems. The encoded  
CC polypeptides can be used to prevent tumour angiogenesis, metastasis and  
CC invasion, and to intervene with pathologies associated with impaired  
CC heparin-binding growth factors, cellular responses to heparin-binding  
CC growth factors and cytokines, cell interaction with plasma lipoproteins,  
CC cellular susceptibility to viral, protozoa and bacterial infections or  
CC disintegration of neurodegenerative plaques. The present sequence  
CC represents a human heparanase protein sequence used in similarity studies  
XX  
XX Sequence 543 AA;  
SQ

Query Match 100.0%; Score 86; DB 5; Length 543;  
Best Local Similarity 100.0%; Pred. No. 8.8e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14  
Db |||||

RESULT 38  
ADD18950  
ID ADD18950 standard; protein; 543 AA.  
XX  
AC ADD18950;  
XX  
XX 15-JAN-2004 (first entry)  
XX Human disease related protein SeqID439.  
XX human; disease state; cytostatic; antiinflammatory; ophthalmological;  
KW antiarteriosclerotic; vulnary; gene therapy;  
KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;  
KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;  
KW glucose transportation; catecholamine synthesis; iron transport;  
KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;  
KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;  
KW inflammatory condition; wound healing.  
XX  
XX Homo sapiens.  
XX WO2003018621-A2.  
XX  
XX 06-MAR-2003.  
XX  
XX 23-AUG-2002; 2002WO-GB003892.  
XX  
XX 23-AUG-2001; 2001GB-00020558.  
PR 05-OCT-2001; 2001GB-00024037.  
XX  
XX (OXFO-) OXFORD BIOMEDICA UK LTD.  
XX  
XX Kingman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;  
PI WPI; 2003-290046/28.  
XX N-PSDB; ADD18951.  
XX  
XX New substantially purified polypeptide, useful for diagnosing or treating

PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion  
PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or  
XX wound healing.  
XX Claim 25; SEQ ID NO 439; 424pp; English.  
XX  
XX This invention relates to novel human genes and gene product which are  
CC implicated in certain disease states. Compounds which modulate the  
CC proteins of the invention may have cytostatic, antiinflammatory,  
CC ophthalmological, antiarteriosclerotic or vulnerary activities. The  
CC sequences of the invention may be useful for gene therapy. The invention  
CC may be useful for diagnosing or treating a hypoxia-regulated condition,  
CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,  
CC erythropoiesis, or the biological response to hypoxia conditions  
CC including processes such as glycolysis, gluconeogenesis, glucose  
CC transportation, catecholamine synthesis, iron transport or nitric oxide  
CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion  
CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,  
CC inflammatory conditions or wound healing. The present sequence is that of  
XX a disease related protein of the invention.  
SQ Sequence 543 AA;  
Query Match 100.0%; Score 86; DB 7; Length 543;  
Best Local Similarity 100.0%; Pred. No. 8.8e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14  
Db |||||

RESULT 39  
ADG88800  
ID ADG88800 standard; protein; 543 AA.  
XX  
AC ADG88800;  
XX  
XX 11-MAR-2004 (first entry)  
XX Human hpa protein.  
XX  
XX Wound healing; heparanase; ulcer; burn; laceration; surgical incision;  
KW necrosis; pressure wound; diabetic ulcer; angiogenesis; human; therapy.  
XX  
XX Homo sapiens.  
XX US2003161823-A1.  
XX  
XX 28-AUG-2003.  
XX  
XX 14-JAN-2003; 2003US-00341582.  
XX  
XX 31-AUG-1998; 98WO-US017954.  
PR 01-MAR-1999; 99US-00258892.  
PR 06-FEB-2001; 2001US-00776874.  
PR 05-SEP-2001; 2001WO-IL000830.  
PR 19-NOV-2001; 2001US-00988113.  
XX  
XX (ILAN/) ILAN N.  
PA (VLOD/) VLODAVSKY I.  
PA (YACO/) YACOBY-ZEEVI O.  
PA (PECK/) PECKER I.  
PA (FEIN/) FEINSTEIN E.  
XX  
XX Ilan N, Vlodavsky I, Yacoby-Zeevi O, Pecker I, Feinstein E;  
XX WPI; 2003-897910/82.  
XX N-PSDB; ADG88799, ADG88801, ADG88832.  
XX  
XX Composition for treating a wound comprising recombinant heparanase is  
PT useful to induce or accelerate wound healing and induce or accelerate  
PT angiogenesis.

XX PS Claim 2; SEQ ID NO 10; 143pp; English.

XX CC The present invention relates to methods and compositions for inducing

XX CC and/or accelerating wound healing via the catalytic activity of

XX CC heparanase. The invention is used to induce or accelerate a healing

XX CC process, particularly of an ulcer, burn, laceration, surgical incision,

XX CC necrosis, pressure wound, diabetic ulcer and to induce or accelerate

XX CC angiogenesis. The present sequence is human hpa protein.

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 7; Length 543;

Best Local Similarity 100.0%; Pred. No. 8.8e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHYYLNGRTATR 14

Db 294 TWHYYLNGRTATR 307

|||||

RESULT 40

ADL16379

ID ADL16379 standard; protein; 543 AA.

XX AC ADL16379;

XX DT 06-MAY-2004 (first entry)

XX DE Human heparanase partial protein.

XX KW Human; heparanase; heparanase-dependent cancer; cancer;

XX KW autoimmune reaction; inflammation; chromosome 4; enzyme.

XX OS Homo sapiens.

XX US2003236215-A1.

XX PD 25-DEC-2003.

XX PF 09-JUN-2003; 2003US-00456573.

XX PR 31-AUG-1998; 98WO-US017954.

XX PR 01-MAR-1999; 99US-00258892.

XX PR 08-NOV-1999; 99US-00435739.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX Pecker I, Vlodavsky I, Feinstein E;

XX WPI; 2004-070610/07.

XX New antisense oligonucleotide hybridizable with a polynucleotide encoding

XX a polypeptide with heparanase activity, useful for treating diseases such

XX as cancer and autoimmune disorders.

XX Claim 3; SEQ ID NO 10; 108pp; English.

XX The invention relates to an antisense oligonucleotide (ASO) comprising a

XX polynucleotide or a polynucleotide analogue of at least 10 bases being

XX hybridizable in vivo, under physiological conditions, with a portion of

XX a polynucleotide strand encoding a polypeptide having heparanase

XX catalytic activity. Also included are a method of in vivo downregulating

XX heparanase activity (comprising administering the ASO in vivo), a method

XX of treating a subject suffering from a pathological condition

XX (characterised by heparanase activity, comprising administering ASO to

XX the subject), a pharmaceutical composition comprising the ASO and a

XX carrier, an antisense nucleic acid construct (comprising a promoter

XX sequence and a polynucleotide sequence directing the synthesis of an

XX antisense RNA sequence of at least 10 bases being hybridizable in vivo,

XX under physiological conditions, with a polynucleotide strand encoding a

XX polypeptide having heparanase catalytic activity), a method of in vivo

CC downregulating heparanase activity (comprising administering in vivo the

CC antisense nucleic acid construct), a pharmaceutical composition

CC comprising the antisense nucleic acid construct and a carrier, and an

CC antisense oligonucleotide comprising a polynucleotide or a polynucleotide

CC analogue of at least 10 bases being hybridisable in vivo, under

CC physiological conditions, with a portion of a polynucleotide strand being

CC characterised by forming at least a portion of an untranslated region

CC (UTR) for a polynucleotide strand encoding a polypeptide having

CC heparanase catalytic activity. The methods and compositions of the

CC present invention are useful for the prevention and/or treatment of

CC diseases or conditions associated with aberrant heparanase activity, such

CC as heparanase-dependent cancer, cancer, autoimmune reaction and

CC inflammation. The gene for human heparanase is located on chromosome 4.

XX The present sequence is a human heparanase protein.

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 8; Length 543;

Best Local Similarity 100.0%; Pred. No. 8.8e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHYYLNGRTATR 14

Db 294 TWHYYLNGRTATR 307

|||||

RESULT 41

ADK52086

ID ADK52086 standard; protein; 543 AA.

XX AC ADK52086;

XX DT 20-MAY-2004 (first entry)

XX DE Human atopic dermatitis/psoriasis-associated protein #1.

XX KW Human; atopic dermatitis; psoriasis; dermatological; anti-inflammatory;

XX KW antipsoriatic; rash.

XX OS Homo sapiens.

XX WO2004016785-A1.

XX PD 26-FEB-2004.

XX PF 06-AUG-2003; 2003WO-JP0099999.

XX PR 06-AUG-2002; 2002JP-00229319.

XX PR 14-MAY-2003; 2003JP-00136544.

XX (GENO-) GENOX RES INC.

XX (UYJU-) UNIV JUNTENDO.

XX Itoh M, Ogawa K, Shinagawa A, Sudo H, Ogawa H, Ra C;

XX Mitsuishi K;

XX WPI; 2004-214514/20.

XX N-PSDB; ADK51968.

XX Detecting atopic dermatitis or psoriasis comprises assaying levels of

XX expression of an indicator gene at a rash site and non-rash site of a

XX person with atopic dermatitis or psoriasis.

XX Example 2; SEQ ID NO 119; 484pp; Japanese.

XX The invention relates to detecting atopic dermatitis or psoriasis

XX comprising assaying the levels of expression of an indicator gene at a

XX rash site and non-rash site of a person with atopic dermatitis or

XX psoriasis, comparing these levels with those of a healthy person, and

XX determining that if the levels of indicators are higher or lower, then

XX this indicates the disease. Also included are a reagent for detecting

XX atopic dermatitis or psoriasis, a kit for screening for treatments, a

XX transgenic non human vertebrate animal models for the diseases, an agent

CC for inducing the diseases in mice and a DNA chip for assaying for the  
CC indicator genes. The method is used for treatment, detection and animal  
CC models for research of atopic dermatitis and psoriasis. The present  
CC sequence is a protein encoded by an indicator gene of the invention.  
XX  
SQ Sequence 543 AA;  
  
Query Match 100.0%; Score 86; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 8.8e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 TWHYYLNGRTATR 14  
|||||  
Db 294 TWHYYLNGRTATR 307  
  
RESULT 42  
ADM48716  
ID ADM48716 standard; protein; 543 AA.  
XX  
AC ADM48716;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE Human hpa protein #1.  
XX  
KW Transgenic animal; heparanase; cancer; viral infection; restenosis;  
KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;  
KW human.  
XX  
OS Homo sapiens.  
XX  
PN US2003217375-A1.  
XX  
PD 20-NOV-2003.  
XX  
PF 24-FEB-2003; 2003US-00371218.  
XX  
PR 31-AUG-1998; 98WO-US017954.  
PR 01-MAR-1999; 99US-00258892.  
PR 06-FEB-2001; 2001US-00776874.  
PR 19-NOV-2001; 2001US-00988113.  
XX  
PA (ZCHA/) ZCHARIA E.  
PA (VLOD/) VLODAVSKY I.  
PA (METZ/) METZGER S.  
PA (PECK/) PECKER I.  
PA (ILAN/) ILAN N.  
PA (CHAJ/) CHAJEK-SHAUL T.  
PA (GOLD/) GOLDSHMIDT O.  
XX  
PI Zcharia E, Vlodayevsky I, Metzger S, Pecker I, Ilan N;  
PI Chajek-Shaul T, Goldshmidt O;  
XX  
WPI; 2004-021918/02.  
DR N-PSDB; ADM48715, ADM48717.  
XX  
PT New transgenic non-human animal expressing heparinase, useful as models  
PT for human disease, such as cancers, viral infection, neurodegenerative  
PT diseases, restenosis, atherosclerosis and pulmonary disorders.  
XX  
PS Example 1; SEQ ID NO 10; 106pp; English.  
XX  
CC The present invention relates to a transgenic non-human animal whose  
CC genome comprises an exogenous polynucleotide sequence, including a  
CC promoter active in tissues of the non-human, a region encoding a human  
CC heparanase, where the promoter and the region encoding human heparanase  
CC are operably linked in the exogenous polynucleotide such that human  
CC heparanase is expressed in at least a portion of the cells of the non-  
CC human animal. The methods and compositions of the present invention are  
CC useful for the production of transgenic animals expressing heparanase, to  
CC be used as models for human diseases such as cancers, viral infection,  
CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary  
CC disorders.

CC disorders. The present sequence is human hpa protein used in the  
CC exemplification of the invention.  
XX  
SQ Sequence 543 AA;  
  
Query Match 100.0%; Score 86; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 8.8e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 TWHYYLNGRTATR 14  
|||||  
Db 294 TWHYYLNGRTATR 307  
  
RESULT 43  
ADM48759  
ID ADM48759 standard; protein; 543 AA.  
XX  
AC ADM48759;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE Human hpa protein #2.  
XX  
KW Transgenic animal; heparanase; cancer; viral infection; restenosis;  
KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;  
KW human.  
XX  
OS Homo sapiens.  
XX  
PN US2003217375-A1.  
XX  
PD 20-NOV-2003.  
XX  
PF 24-FEB-2003; 2003US-00371218.  
XX  
PR 31-AUG-1998; 98WO-US017954.  
PR 01-MAR-1999; 99US-00258892.  
PR 06-FEB-2001; 2001US-00776874.  
PR 19-NOV-2001; 2001US-00988113.  
XX  
PA (ZCHA/) ZCHARIA E.  
PA (VLOD/) VLODAVSKY I.  
PA (METZ/) METZGER S.  
PA (PECK/) PECKER I.  
PA (ILAN/) ILAN N.  
PA (CHAJ/) CHAJEK-SHAUL T.  
PA (GOLD/) GOLDSHMIDT O.  
XX  
PI Zcharia E, Vlodayevsky I, Metzger S, Pecker I, Ilan N;  
PI Chajek-Shaul T, Goldshmidt O;  
XX  
WPI; 2004-021918/02.  
DR N-PSDB; ADM48748.  
XX  
PT New transgenic non-human animal expressing heparinase, useful as models  
PT for human disease, such as cancers, viral infection, neurodegenerative  
PT diseases, restenosis, atherosclerosis and pulmonary disorders.  
XX  
PS Example 10; Fig 16; 106pp; English.  
XX  
CC The present invention relates to a transgenic non-human animal whose  
CC genome comprises an exogenous polynucleotide sequence, including a  
CC promoter active in tissues of the non-human, a region encoding a human  
CC heparanase, where the promoter and the region encoding human heparanase  
CC are operably linked in the exogenous polynucleotide such that human  
CC heparanase is expressed in at least a portion of the cells of the non-  
CC human animal. The methods and compositions of the present invention are  
CC useful for the production of transgenic animals expressing heparanase, to  
CC be used as models for human diseases such as cancers, viral infection,  
CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary  
CC disorders. The present sequence is human hpa protein used in the  
CC exemplification of the invention.

```

XX SQ Sequence 543 AA;
Query Match 100.0%; Score 86; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
    |||
Db 294 TWHYYLNGRTATR 307

RESULT 44
ADN05074
ID ADN05074 standard; protein; 543 AA.
XX AC ADN05074;
XX DT 01-JUL-2004 (first entry)
XX DE Antipsoriatic protein sequence #716.
XX KW antipsoriatic; gene therapy; psoriasis; diagnosis.
XX OS Homo sapiens.
XX PN WO2004028479-A2.
XX PD 08-APR-2004.
XX PF 25-SEP-2003; 2003WO-US030907.
XX PR 25-SEP-2002; 2002US-0414006P.
XX PA (GETH ) GENENTECH INC.
XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
XX PI Wu TD;
XX DR WPI; 2004-305105/28.
XX DR N-PSDB; ADN04901.
XX PT New PRO nucleic acid or polypeptide, useful for preparing a
XX PT pharmaceutical composition for diagnosing or treating psoriasis in a
XX OS Homo sapiens.
XX PN WO2004028479-A2.
XX PD 08-APR-2004.
XX PF 25-SEP-2003; 2003WO-US030907.
XX PR 25-SEP-2002; 2002US-0414006P.
XX PA (GETH ) GENENTECH INC.
XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
XX PI Wu TD;
XX DR WPI; 2004-305105/28.
XX DR N-PSDB; ADN05073.
XX PT New PRO nucleic acid or polypeptide, useful for preparing a
XX PT pharmaceutical composition for diagnosing or treating psoriasis in a
XX PT mammal.
XX PS Claim 9; SEQ ID NO 1468; 3069pp; English.
XX CC The invention relates to novel polynucleotide and polypeptides for
XX CC treating psoriasis or a sequence having at least 80% identity to the
XX CC above sequences. The nucleic acid is useful for preparing a composition
XX CC for diagnosing or treating psoriasis in a mammal. This sequence
XX CC corresponds to one of the polypeptides of the invention.
XX SQ Sequence 543 AA;
Query Match 100.0%; Score 86; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
    |||
Db 294 TWHYYLNGRTATR 307

RESULT 45
ADN04902
ID ADN04902 standard; protein; 543 AA.
XX AC ADN04902;
XX DT 01-JUL-2004 (first entry)
XX DE Antipsoriatic protein sequence #631.
XX KW antipsoriatic; gene therapy; psoriasis; diagnosis.
XX OS Homo sapiens.
XX PN WO2004028479-A2.
XX PD 08-APR-2004.
XX PF 25-SEP-2003; 2003WO-US030907.
XX PR 25-SEP-2002; 2002US-0414006P.
XX PA (GETH ) GENENTECH INC.
XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
XX PI Wu TD;
XX DR WPI; 2004-305105/28.
XX DR N-PSDB; ADN04901.
XX PT New PRO nucleic acid or polypeptide, useful for preparing a
XX PT pharmaceutical composition for diagnosing or treating psoriasis in a
XX PT mammal.
XX PS Claim 9; SEQ ID NO 1296; 3069pp; English.
XX CC The invention relates to novel polynucleotide and polypeptides for
XX CC treating psoriasis or a sequence having at least 80% identity to the
XX CC above sequences. The nucleic acid is useful for preparing a composition
XX CC for diagnosing or treating psoriasis in a mammal. This sequence
XX CC corresponds to one of the polypeptides of the invention.
XX SQ Sequence 543 AA;
Query Match 100.0%; Score 86; DB 8; Length 543;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
    |||
Db 294 TWHYYLNGRTATR 307

RESULT 46
ADN063831
ID ADN063831 standard; protein; 543 AA.
XX AC ADN063831;
XX DT 26-AUG-2004 (first entry)
XX DE Human heparanase mutant E378A.
XX KW Human; heparanase; heparanase-derived protein; heparanase mutant;
XX KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;
XX KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;
XX KW vascular graft; platelet adhesion; platelet aggregation;
XX KW adhesion disorder; IAD; leukocyte adhesion deficiency;
XX KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;
XX KW vulnery; mutant; mutein; enzyme.
XX OS Homo sapiens.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT Active-site /note= "Active site proton donor"
XX FT Active-site /note= "Active site nucleophile"
XX FT Misc-difference /note= "Ala replaces wild-type Glu"

```

XX WO2004048558-A2.  
 XX 10-JUN-2004.  
 XX 24-NOV-2003; 2003WO-IL000989.  
 XX 24-NOV-2002; 2002IL-00153059.  
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
 XX WPI; 2004-450373/42.  
 XX New nucleic acid construct comprising heparanase-derived polypeptide,  
 XX useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 XX thrombasthenia, or Bernard-Soulier syndrome.  
 XX Example 4; Page; 128pp; English.  
 XX The invention relates to nucleic acid constructs comprising a nucleic  
 XX acid encoding a heparanase-derived protein which lacks heparanase  
 XX endoglycosidase catalytic activity but which retains its cell-cell and  
 XX cell-matrix adhesion properties. The constructs of the invention  
 XX optionally further comprise operably linked regulatory elements. The  
 XX invention also relates to the heparanase-derived proteins and host cells  
 XX comprising the nucleic acid constructs of the invention. The heparanase-  
 XX derived proteins are especially mutants of human heparanase in which the  
 XX active site proton donor Glu225 and/or the active site nucleophile Glu343  
 XX are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 XX optionally further comprise an avian heparanase signal peptide (ADO63825-  
 XX ADO63827). The heparanase-derived protein, nucleic acid construct and  
 XX host cells are useful in preparing a tissue sealant composition for  
 XX sealing injuries, reducing the loss of blood, accelerating the healing  
 XX and homeostasis of an injury, accelerating blood vessel endothelium  
 XX formation or the endothelialisation of vascular grafts, accelerating the  
 XX adhesive activity of mammalian cells, and accelerating the adhesion and  
 XX aggregation of platelets. They may also be used in the treatment of  
 XX disorders associated with adhesion deficiency such as LAD (leukocyte  
 XX adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 XX function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 XX cells of the invention may additionally be used to screen for modulators of  
 XX cell-cell and cell-matrix adhesion, and to prepare an implantable  
 XX synthetic vascular graft comprising a tube made of a biocompatible  
 XX material lined with the cells. The present sequence represents a human  
 XX heparanase mutant E378A created in an example of the invention which  
 XX retains its heparanase catalytic activity. The present sequence is not  
 XX shown in the invention, but is derived from the protein sequence of  
 XX GenBank accession number AF144325 and the information provided on page  
 XX 70.  
 XX Sequence 543 AA;  
 XX Query Match 100.0%; Score 86; DB 8; Length 543;  
 XX Best Local Similarity 100.0%; Pred. No. 8.8e-06;  
 XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TWHYYLNGRTATR 14  
 Db 294 TWHYYLNGRTATR 307  
 RESULT 47  
 ADO63824  
 ID ADO63824 standard; protein; 543 AA.  
 XX ADO63824;  
 XX 26-AUG-2004 (first entry)  
 XX Human heparanase mutant E225A/E343A, SEQ ID:9.  
 XX

KW Human; heparanase; heparanase-derived protein; heparanase mutant;  
 KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;  
 KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;  
 KW vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnerability; mutant; mutein.  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 XX Misc-difference 225 /note= "Ala replaces wild-type Glu (active site proton  
 XX donor)"  
 XX Misc-difference 343 /note= "Ala replaces wild-type Glu (active site  
 XX nucleophile)"  
 XX WO2004048558-A2.  
 XX 10-JUN-2004.  
 XX 24-NOV-2003; 2003WO-IL000989.  
 XX 24-NOV-2002; 2002IL-00153059.  
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
 XX WPI; 2004-450373/42.  
 XX N-PSDB; ADO63818.  
 XX New nucleic acid construct comprising heparanase-derived polypeptide,  
 XX useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 XX thrombasthenia, or Bernard-Soulier syndrome.  
 XX Claim 9; SEQ ID NO 9; 128pp; English.  
 XX The invention relates to nucleic acid constructs comprising a nucleic  
 XX acid encoding a heparanase-derived protein which lacks heparanase  
 XX endoglycosidase catalytic activity but which retains its cell-cell and  
 XX cell-matrix adhesion properties. The constructs of the invention  
 XX optionally further comprise operably linked regulatory elements. The  
 XX invention also relates to the heparanase-derived proteins and host cells  
 XX comprising the nucleic acid constructs of the invention. The heparanase-  
 XX derived proteins are especially mutants of human heparanase in which the  
 XX active site proton donor Glu225 and/or the active site nucleophile Glu343  
 XX are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 XX optionally further comprise an avian heparanase signal peptide (ADO63825-  
 XX ADO63827). The heparanase-derived protein, nucleic acid construct and  
 XX host cells are useful in preparing a tissue sealant composition for  
 XX sealing injuries, reducing the loss of blood, accelerating the healing  
 XX and homeostasis of an injury, accelerating blood vessel endothelium  
 XX formation or the endothelialisation of vascular grafts, accelerating the  
 XX adhesive activity of mammalian cells, and accelerating the adhesion and  
 XX aggregation of platelets. They may also be used in the treatment of  
 XX disorders associated with adhesion deficiency such as LAD (leukocyte  
 XX adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 XX function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 XX cells of the invention may additionally be used to screen for modulators of  
 XX cell-cell and cell-matrix adhesion, and to prepare an implantable  
 XX synthetic vascular graft comprising a tube made of a biocompatible  
 XX material lined with the cells. The present sequence represents the human  
 XX heparanase double mutant E225A/E343A.  
 XX Sequence 543 AA;  
 XX Query Match 100.0%; Score 86; DB 8; Length 543;  
 XX Best Local Similarity 100.0%; Pred. No. 8.8e-06;  
 XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 1 TWHYYLNGRTATR 14  
 DB 294 TWHYYLNGRTATR 307

RESULT 48  
 ADO63823  
 ID ADO63823 standard; protein; 543 AA.  
 XX  
 AC ADO63823;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Human heparanase mutant E343A, SEQ ID:8.  
 XX  
 KW Human; heparanase; heparanase-derived protein; heparanase mutant;  
 KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;  
 KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;  
 KW vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnery; mutant; mutein.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Active-site 225 /note= "Active site proton donor"  
 FT Misc-difference 343 /note= "Ala replaces wild-type Glu (active site  
 FT nucleophile)"  
 FT  
 XX  
 PN WO2004048558-A2.  
 PN  
 PD 10-JUN-2004.  
 PD  
 XX  
 XX 24-NOV-2003; 2003WO-IL000989.  
 XX  
 XX 24-NOV-2002; 2002IL-00153059.  
 XX  
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX  
 PA Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
 PI WPI; 2004-450373/42.  
 DR N-PSDB; ADO63817.  
 DR  
 XX  
 XX New nucleic acid construct comprising heparanase-derived polypeptide,  
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 PT thrombasthenia, or Bernard-Soulier syndrome.  
 PT  
 XX  
 PS Claim 9; SEQ ID NO 8; 128pp; English.  
 XX  
 CC The invention relates to nucleic acid constructs comprising a nucleic  
 CC acid encoding a heparanase-derived protein which lacks heparanase  
 CC endoglycosidase catalytic activity but which retains its cell-cell and  
 CC cell-matrix adhesion properties. The constructs of the invention  
 CC optionally further comprise operably linked regulatory elements. The  
 CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the  
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
 CC host cells are useful in preparing a tissue sealant composition for  
 CC sealing injuries, reducing the loss of blood, accelerating the healing  
 CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialisation of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be use in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte

CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents the human  
 CC heparanase mutant E343A.  
 XX  
 SQ Sequence 543 AA;  
 XX  
 Query Match 100.0%; Score 86; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 TWHYYLNGRTATR 14  
 DB 294 TWHYYLNGRTATR 307

RESULT 49  
 ADO63832  
 ID ADO63832 standard; protein; 543 AA.  
 XX  
 AC ADO63832;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Human heparanase mutant E396A.  
 XX  
 KW Human; heparanase; heparanase-derived protein; heparanase mutant;  
 KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;  
 KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;  
 KW vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnery; mutant; mutein; enzyme.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Active-site 225 /note= "Active site proton donor"  
 FT Active-site 343 /note= "Active site nucleophile"  
 FT Misc-difference 396 /note= "Ala replaces wild-type Glu"  
 FT  
 XX  
 PN WO2004048558-A2.  
 PN  
 PD 10-JUN-2004.  
 PD  
 XX  
 XX 24-NOV-2003; 2003WO-IL000989.  
 XX  
 XX 24-NOV-2002; 2002IL-00153059.  
 XX  
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX  
 PA Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
 PI WPI; 2004-450373/42.  
 DR N-PSDB; ADO63817.  
 DR  
 XX  
 XX New nucleic acid construct comprising heparanase-derived polypeptide,  
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 PT thrombasthenia, or Bernard-Soulier syndrome.  
 PT  
 XX  
 PS Example 4; Page; 128pp; English.  
 XX  
 CC The invention relates to nucleic acid constructs comprising a nucleic  
 CC acid encoding a heparanase-derived protein which lacks heparanase  
 CC endoglycosidase catalytic activity but which retains its cell-cell and  
 CC cell-matrix adhesion properties. The constructs of the invention  
 CC optionally further comprise operably linked regulatory elements. The  
 CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the  
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
 CC host cells are useful in preparing a tissue sealant composition for  
 CC sealing injuries, reducing the loss of blood, accelerating the healing  
 CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialisation of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be use in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte

CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the  
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
 CC host cells are useful in preparing a tissue sealant composition for  
 CC sealing injuries, reducing the loss of blood, accelerating the healing  
 CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialisation of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be used in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte  
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be used to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents a human  
 CC heparanase mutant E378A created in an example of the invention which  
 CC retains its heparanase catalytic activity. The present sequence is not  
 CC shown in the invention, but is derived from the protein sequence of  
 CC GenBank accession number AF144325 and the information provided on page  
 CC 70.

XX  
 SQ Sequence 543 AA;

Query Match 100.0%; Score 86; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHYYLNGRTATR 14  
 |||||  
 Db 294 TWHYYLNGRTATR 307

RESULT 50

ADO63822  
 ID ADO63822 standard; protein; 543 AA.

XX ADO63822;

DT 26-AUG-2004 (first entry)

DE Human heparanase mutant E225A, SEQ ID: 7.

XX Human; heparanase; heparanase-derived protein; heparanase mutant;  
 KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;  
 KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;  
 KW vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulneryary; mutant; mutein.

OS Homo sapiens.  
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 225  
 /note= "Ala replaces wild-type Glu (active site proton  
 donor)"

FT Active-site 343  
 /note= "Active site nucleophile"

XX WO2004048558-A2.

XX 10-JUN-2004.

XX 24-NOV-2003; 2003WO-11000989.

XX 24-NOV-2002; 2002IL-00153059.

XX

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;

XX WPI; 2004-450373/42.

XX N-PSDB; ADO63816.

XX New nucleic acid construct comprising heparanase-derived polypeptide,  
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 PT thrombasthenia, or Bernard-Soulier syndrome.

XX Claim 9; SEQ ID NO 7; 128pp; English.

XX The invention relates to nucleic acid constructs comprising a nucleic  
 CC acid encoding a heparanase-derived protein which lacks heparanase  
 CC endoglycosidase catalytic activity but which retains its cell-cell and  
 CC cell-matrix adhesion properties. The constructs of the invention  
 CC optionally further comprise operably linked regulatory elements. The  
 CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the  
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
 CC host cells are useful in preparing a tissue sealant composition for  
 CC sealing injuries, reducing the loss of blood, accelerating the healing  
 CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialisation of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be used in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte  
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be used to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents the human  
 CC heparanase mutant E225A.

XX Sequence 543 AA;

Query Match 100.0%; Score 86; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWHYYLNGRTATR 14  
 |||||  
 Db 294 TWHYYLNGRTATR 307

Search completed: June 5, 2006, 12:42:27

Job time : 121.425 secs



103	38	44.2	691	2	JE0150	acetylcholinestera	176	36.5	42.4	192	2	S42999	viral infectivity
104	38	44.2	746	2	A25363	acetylcholinestera	177	36.5	42.4	192	2	S42991	viral infectivity
105	37.5	43.6	192	2	B53294	superoxide dismuta	178	36.5	42.4	192	2	S42949	viral infectivity
106	37.5	43.6	199	2	T50046	superoxide dismuta	179	36.5	42.4	192	2	S42993	viral infectivity
107	37.5	43.6	200	2	T50045	superoxide dismuta	180	36.5	42.4	200	1	A34319	superoxide dismuta
108	37.5	43.6	546	2	JC4798	seizure-related me	181	36.5	42.4	200	2	H84274	superoxide dismuta
109	37.5	43.6	550	2	I37579	DiGeorge syndrome	182	36.5	42.4	200	2	T50043	superoxide dismuta
110	37	43.0	86	1	D46335	W protein - Maedi/	183	36.5	42.4	203	2	T50044	superoxide dismuta
111	37	43.0	128	2	S63983	bile acid-binding	184	36.5	42.4	244	2	T50936	dehydrogenase/relu
112	37	43.0	129	2	G81206	hypothetical prote	185	36.5	42.4	315	2	T29045	hypothetical prote
113	37	43.0	150	2	T39482	N-acetyltransferas	186	36.5	42.4	497	2	H83886	hypothetical prote
114	37	43.0	155	2	F83341	hypothetical prote	187	36.5	42.4	663	2	E83623	hypothetical prote
115	37	43.0	157	2	E83761	hypothetical prote	188	36.5	42.4	1254	1	JQ1978	structural polypro
116	37	43.0	181	2	A36367	ADP-ribosylation f	189	36.5	42.4	1254	1	VHVVVE	structural polypro
117	37	43.0	181	2	A36367	ADP-ribosylation f	190	36.5	42.4	1254	1	VHVVVE	structural polypro
118	37	43.0	191	2	S29008	ADP-ribosylation f	191	36.5	42.4	1254	1	VHVVVT	structural polypro
119	37	43.0	224	2	B81783	hypothetical prote	192	36.5	42.4	1255	1	B44213	structural polypro
120	37	43.0	225	2	B81207	hypothetical prote	193	36.5	42.4	1255	1	D44213	structural polypro
121	37	43.0	277	2	S46330	hypothetical prote	194	36.5	42.4	2100	2	T03223	probable polyketid
122	37	43.0	292	2	S71556	DNA-binding protei	195	36	41.9	94	2	B82518	hypothetical prote
123	37	43.0	297	2	F83433	DNA-3-methyladenin	196	36	41.9	174	2	D83692	hypothetical prote
124	37	43.0	361	2	F72862	hypothetical prote	197	36	41.9	184	2	F84904	hypothetical prote
125	37	43.0	362	2	T41842	AcMNPV orf101 - Bo	198	36	41.9	204	2	S64838	hypothetical prote
126	37	43.0	369	2	G69254	golicolol-P-glucose	199	36	41.9	235	2	A81174	hypothetical prote
127	37	43.0	386	2	T12845	hypothetical prote	200	36	41.9	244	2	C75605	hypothetical prote
128	37	43.0	405	2	D83351	hypothetical prote	201	36	41.9	247	2	G64341	hypothetical prote
129	37	43.0	405	2	G75027	alanyl-tRNA synthe	202	36	41.9	250	2	S07237	hypothetical prote
130	37	43.0	448	2	C95257	choline binding pr	203	36	41.9	255	2	A60944	ubiquinol-cytochro
131	37	43.0	448	2	C98122	choline binding pr	204	36	41.9	281	2	I64249	hypothetical prote
132	37	43.0	458	2	A46366	galactokinase (8C	205	36	41.9	282	2	Ar2329	tRNA-pseudouridine
133	37	43.0	484	2	F71061	hypothetical prote	206	36	41.9	288	2	D64694	hypothetical prote
134	37	43.0	494	2	AC3582	probable blue-copp	207	36	41.9	301	2	E95977	UTP-glucose-1-phos
135	37	43.0	517	1	S19243	tyrosinase-related	208	36	41.9	307	2	AE3054	UTP-glucose-1-phos
136	37	43.0	519	1	YRHUR2	dopachrome isomera	209	36	41.9	307	2	H98231	exon protein (impo
137	37	43.0	522	2	I51245	tyrosinase related	210	36	41.9	311	2	S66011	transcription regu
138	37	43.0	533	2	S51171	amino acid transpo	211	36	41.9	316	2	T19884	hypothetical prote
139	37	43.0	537	1	YRHUB6	tyrosinase-related	212	36	41.9	320	1	D49349	UTP-glucose-1-phos
140	37	43.0	537	1	YRMSB6	tyrosinase-related	213	36	41.9	332	2	T14624	hypothetical prote
141	37	43.0	546	1	C70393	probable adenyl-yl-	214	36	41.9	332	2	T45723	hypothetical prote
142	37	43.0	548	2	T15318	hypothetical prote	215	36	41.9	336	2	C71964	UDP-3-O-[3-hydroxy
143	37	43.0	552	2	E72283	alpha-galactosidas	216	36	41.9	336	2	D64544	UDP-3-O-[3-hydroxy
144	37	43.0	557	2	T27752	hypothetical prote	217	36	41.9	353	2	D69422	F420-nonreducing h
145	37	43.0	593	2	JCS167	protein-tyrosine-p	218	36	41.9	356	1	S64902	probable sugar red
146	37	43.0	610	2	F95898	probable nodulatio	219	36	41.9	380	2	T18509	hypothetical prote
147	37	43.0	676	2	S41217	hypothetical prote	220	36	41.9	380	2	T21112	hypothetical prote
148	37	43.0	751	2	A81816	nitric oxide reduc	221	36	41.9	392	2	T24240	hypothetical prote
149	37	43.0	751	2	D81062	nitric oxide reduc	222	36	41.9	421	2	C96806	unknown protein T5
150	37	43.0	801	2	A47744	diacylglycerol kin	223	36	41.9	423	2	T19145	hypothetical prote
151	37	43.0	818	2	T29560	hypothetical prote	224	36	41.9	431	2	E81357	glutamate-tRNA lig
152	37	43.0	927	2	T38518	ribonuclease II RN	225	36	41.9	441	2	JQ2191	nucleocapsid prote
153	37	43.0	946	2	F84280	ATP-dependent heli	226	36	41.9	463	1	D70585	probable glycs prot
154	37	43.0	1024	2	G72041	exodeoxyribonuclea	227	36	41.9	480	2	C96744	hypothetical prote
155	37	43.0	1024	2	F86582	exodeoxyribonuclea	228	36	41.9	502	2	T49188	serin carboxypepti
156	37	43.0	1024	2	D81624	exodeoxyribonuclea	229	36	41.9	504	2	G87532	tryptophan halogen
157	37	43.0	1032	2	F65071	hypothetical prote	230	36	41.9	512	2	A72866	major budded virus
158	37	43.0	1032	2	C85943	probable oxidoredu	231	36	41.9	529	1	VGNVAC	major envelope gly
159	37	43.0	1032	2	G91097	probable oxidoredu	232	36	41.9	530	2	T41865	GP64/67 EFP orf128
160	37	43.0	1126	2	S04716	DNA-directed RNA p	233	36	41.9	530	2	I40608	capA protein - Clo
161	37	43.0	1156	2	E69444	chromosome segrega	234	36	41.9	627	2	T49952	hypothetical prote
162	37	43.0	1171	2	T00380	KIAA0637 protein -	235	36	41.9	635	2	S57714	csfPB protein - Clo
163	37	43.0	1208	2	AE1947	chromosome segrega	236	36	41.9	639	2	S23118	proprotein convert
164	37	43.0	1386	2	T49316	profilaggrin relat	237	36	41.9	671	2	S61099	leukotriene-A4 hyd
165	37	43.0	2491	1	A28372	insulin-like growt	238	36	41.9	1536	2	S59841	4-alpha-glucanotra
166	37	43.0	2499	1	A30788	mannose 6-phosphat	239	36	41.9	2434	2	S44861	DNA topoisomerase
167	37	43.0	2629	2	T30987	telomerase-associa	240	36	41.9	2787	2	S45416	TEU1 protein - yea
168	37	43.0	2824	2	T22759	hypothetical prote	241	35.5	41.3	50	2	B41308	vif protein - huma
169	36.5	42.4	192	2	S43007	viral infectivity	242	35.5	41.3	185	2	T36546	hypothetical prote
170	36.5	42.4	192	2	S43006	viral infectivity	243	35.5	41.3	192	1	ASLJNA	vif protein - huma
171	36.5	42.4	192	2	S42997	viral infectivity	244	35.5	41.3	192	1	ASLJJS3	vif protein - huma
172	36.5	42.4	192	2	S42958	viral infectivity	245	35.5	41.3	192	2	S42954	viral infectivity
173	36.5	42.4	192	2	S42968	viral infectivity	246	35.5	41.3	192	2	S42964	viral infectivity
174	36.5	42.4	192	2	S42941	viral infectivity	247	35.5	41.3	192	2	S42995	viral infectivity
175	36.5	42.4	192	2	S42992	viral infectivity	248	35.5	41.3	192	2	S42975	viral infectivity

249	35.5	41.3	192	2	S42946	viral infectivity	322	35	40.7	864	2	H85335	hypothetical prote
250	35.5	41.3	271	2	B83027	thiosulfate sulfur	323	35	40.7	864	2	T04518	hypothetical prote
251	35.5	41.3	328	1	Q0BER6	HVLF4 protein - hu	324	35	40.7	900	2	G87431	pyruvate phosphate
252	35.5	41.3	647	2	B82579	acetyl coenzyme A	325	35	40.7	964	2	E71460	probable outer mem
253	35.5	41.3	873	2	JC4863	homeobox protein z	326	35	40.7	1056	2	E96748	hypothetical prote
254	35.5	41.3	1236	1	VHWWE	structural polypro	327	35	40.7	1182	2	T13952	membrane protein p
255	35	40.7	53	2	S39074	light-harvesting p	328	35	40.7	1245	2	E83110	exodeoxyribonuclea
256	35	40.7	107	2	A64643	hypothetical prote	329	35	40.7	1273	2	T00338	hypothetical prote
257	35	40.7	182	2	D83638	conserved hypotet	330	35	40.7	1369	2	T17504	hypothetical prote
258	35	40.7	196	2	T29343	hypothetical prote	331	35	40.7	1397	2	C64805	rhsc protein precu
259	35	40.7	205	2	B84464	hypothetical prote	332	35	40.7	1447	2	G86474	probable protein g
260	35	40.7	212	2	A81190	conserved hypotet	333	35	40.7	1630	2	S64403	ESPl protein - yea
261	35	40.7	212	2	G81913	hypothetical prote	334	35	40.7	1649	2	C96822	hypothetical prote
262	35	40.7	238	2	G90418	ABC transporter, A	335	35	40.7	2090	2	S26058	probable transform
263	35	40.7	249	2	E64404	hypothetical prote	336	35	40.7	3212	2	T24692	hypothetical prote
264	35	40.7	252	2	S77108	hypothetical prote	337	35	40.7	26926	1	I38344	titin, cardiac mus
265	35	40.7	274	2	PRWVM	HIV-1 retropepin	338	34.5	40.1	193	2	S25208	prfJ protein - Esc
266	35	40.7	279	2	JG0164	LIM protein, FHL4	339	34.5	40.1	203	2	A42710	superoxide dismuta
267	35	40.7	295	2	S46413	hypothetical prote	340	34.5	40.1	204	2	D71339	probable ribosomal
268	35	40.7	302	2	H69823	conserved hypotet	341	34.5	40.1	344	1	KHPGD	cathepsin D (EC 3.
269	35	40.7	312	2	B96512	hypothetical prote	342	34.5	40.1	746	2	T24978	hypothetical prote
270	35	40.7	323	2	AC0117	hypothetical prote	343	34.5	40.1	748	2	S66129	disintegrin (EC 3.
271	35	40.7	330	2	H71309	probable asparagin	344	34.5	40.1	825	1	A40026	neurotrophin-3 rec
272	35	40.7	330	2	T49002	hypothetical prote	345	34.5	40.1	958	1	P1BVCC	la protein - cowpe
273	35	40.7	332	2	S60935	hypothetical prote	346	34	39.5	74	2	T44088	probable transposa
274	35	40.7	333	2	A41881	collagenase PTC (	347	34	39.5	118	2	F34792	IG heavy chain pre
275	35	40.7	335	2	T20920	hypothetical prote	348	34	39.5	120	2	S36306	T-cell receptor de
276	35	40.7	340	2	T41729	probable adenosine	349	34	39.5	129	2	A75558	hypothetical prote
277	35	40.7	340	2	E69544	hypothetical prote	350	34	39.5	130	2	AH0463	probable acetyltra
278	35	40.7	351	2	T42538	adenosine kinase h	351	34	39.5	141	2	E87790	protein B0207.8 [i
279	35	40.7	351	2	H71432	probable glucosylt	352	34	39.5	146	2	C90477	quinol oxidase-2,
280	35	40.7	354	2	JQ0413	alkanal monooxygen	353	34	39.5	147	2	T01039	DNA mismatch endon
281	35	40.7	354	2	T33270	hypothetical prote	354	34	39.5	156	2	B85814	DNA mismatch endon
282	35	40.7	372	2	AE3191	conserved hypotet	355	34	39.5	156	2	B90966	patch repair prote
283	35	40.7	385	2	E86359	Similar to seed ma	356	34	39.5	156	2	AG0754	DNA mismatch endon
284	35	40.7	388	2	S04110	Integrase - Strept	357	34	39.5	156	2	JS0264	hypothetical prote
285	35	40.7	389	2	AH0547	methylicitrate synt	358	34	39.5	157	2	H61222	conserved hypotet
286	35	40.7	399	2	B90677	probable citrate s	359	34	39.5	158	2	A12899	hypothetical prote
287	35	40.7	399	2	E64760	citrate (si)-synth	360	34	39.5	158	2	B97675	hypothetical prote
288	35	40.7	399	2	B85527	hypothetical prote	361	34	39.5	178	2	A71730	ubiquinone biosynt
289	35	40.7	399	2	G83722	hypothetical prote	362	34	39.5	179	2	G97729	ubiquinone biosynt
290	35	40.7	393	2	S77691	probable finger pr	363	34	39.5	183	2	S31016	gene 71 protein -
291	35	40.7	438	2	B83017	probable MFS trans	364	34	39.5	184	2	S24450	terminase - phage
292	35	40.7	443	2	G90388	thermopsine precu	365	34	39.5	186	2	G97096	integrase/recombin
293	35	40.7	470	2	S33639	finger protein esc	366	34	39.5	191	2	G64017	hypothetical prote
294	35	40.7	471	2	T33997	hypothetical prote	367	34	39.5	193	2	E64488	hypothetical prote
295	35	40.7	472	2	S50859	P2x receptor - rat	368	34	39.5	196	2	D90389	conserved hypotet
296	35	40.7	477	2	T45722	hypothetical prote	369	34	39.5	198	2	S76039	probable orotate p
297	35	40.7	494	2	B95411	probable aldehyde	370	34	39.5	200	1	A54020	Crotalus neutraliz
298	35	40.7	496	2	S33791	ARS-binding protei	371	34	39.5	205	2	D84527	probable ADP-ribos
299	35	40.7	511	2	C95205	hypothetical prote	372	34	39.5	206	2	AC2443	probable phosphorib
300	35	40.7	512	1	ALBSL	alpha-amylase (EC	373	34	39.5	222	2	G86944	probable membrane
301	35	40.7	527	2	D84517	probable replicati	374	34	39.5	225	2	T34201	hypothetical prote
302	35	40.7	529	2	S46116	probable regulator	375	34	39.5	234	2	S26453	hypothetical prote
303	35	40.7	549	2	S49446	RING-finger protei	376	34	39.5	237	2	S26439	hypothetical prote
304	35	40.7	560	2	AB2437	NADH dehydrogenase	377	34	39.5	239	2	A11997	hypothetical prote
305	35	40.7	570	2	T30527	hypothetical prote	378	34	39.5	240	2	B82383	conserved hypotet
306	35	40.7	587	2	T32546	hypothetical prote	379	34	39.5	245	2	I57946	thiopurine methylt
307	35	40.7	594	2	A82913	hypothetical prote	380	34	39.5	259	2	T36003	hypothetical prote
308	35	40.7	617	2	E72803	gp31 protein - Myc	381	34	39.5	260	2	E83174	hypothetical prote
309	35	40.7	626	2	C70801	hypothetical prote	382	34	39.5	261	2	C97204	thioesterase limpo
310	35	40.7	655	2	B65217	hypothetical 73.7	383	34	39.5	270	2	A71907	outer membrane pro
311	35	40.7	665	2	E86102	hypothetical prote	384	34	39.5	274	2	AH1203	transketolase homo
312	35	40.7	665	2	B91262	hypothetical prote	385	34	39.5	277	2	AE1831	hypothetical prote
313	35	40.7	665	2	G82208	G8DEF family prote	386	34	39.5	278	2	F89044	protein B0238.10 [
314	35	40.7	713	2	F82506	probable TonB syst	387	34	39.5	287	2	D83856	hypothetical prote
315	35	40.7	716	2	D69855	conserved hypotet	388	34	39.5	292	2	F70877	hypothetical prote
316	35	40.7	770	2	A11769	autolysin, amidase	389	34	39.5	293	2	T06232	Ps16 protein - whe
317	35	40.7	780	2	I47038	vasopressin-activa	390	34	39.5	294	2	T05725	cp31AHV protein -
318	35	40.7	785	2	T19741	hypothetical prote	391	34	39.5	297	1	NBR1	apolipoprotein H p
319	35	40.7	802	2	C83588	probable hydroxama	392	34	39.5	303	2	S23780	nucleic acid-bindi
320	35	40.7	809	2	T32899	probable leukotrie	393	34	39.5	306	2	T09067	extensin-like prot
321	35	40.7	846	2	T27282	hypothetical prote	394	34	39.5	310	2	E64751	probable membrane

395	34	39.5	311	2	D82786	thiamin biosynthes	468	34	39.5	647	2	G75060	hydogenase-4 comp
396	34	39.5	316	2	H82649	hypothetical prote	469	34	39.5	648	2	S59723	transcriptional acti
397	34	39.5	318	2	C70636	probable echa13 pr	470	34	39.5	661	2	T38176	hypothetical Myb f
398	34	39.5	328	2	AH1226	N-acetylmuramoyl-L	471	34	39.5	677	2	G69895	formate dehydrogen
399	34	39.5	330	2	G95187	conserved domain p	472	34	39.5	735	2	E87601	methyl-accepting c
400	34	39.5	333	2	A49405	protein kinase Pkn	473	34	39.5	734	2	E69745	hypothetical prote
401	34	39.5	334	2	A82751	ABC transporter At	474	34	39.5	829	2	I40014	sorbitol dehydroge
402	34	39.5	351	2	E84096	hypothetical prote	475	34	39.5	910	2	T29935	hypothetical prote
403	34	39.5	353	2	A86766	hypothetical prote	476	34	39.5	982	2	T19526	hypothetical prote
404	34	39.5	357	2	T23460	hypothetical prote	477	34	39.5	990	2	B49351	bacteriophage N4 a
405	34	39.5	374	2	F97309	uncharacterized co	478	34	39.5	990	2	H90703	bacteriophage N4 a
406	34	39.5	377	2	JC4612	actin - Chlamydomo	479	34	39.5	990	2	C85554	bacteriophage N4 a
407	34	39.5	377	2	SL1420	actin - Volvox car	480	34	39.5	995	2	C81593	polymorphic membra
408	34	39.5	379	2	T31154	hypothetical prote	481	34	39.5	998	2	S31735	NAD ADP-ribosyltra
409	34	39.5	381	2	AI0159	probable periplasm	482	34	39.5	1011	1	JH0581	NAD ADP-ribosyltra
410	34	39.5	383	2	T15698	hypothetical prote	483	34	39.5	1041	2	C83548	hypothetical prote
411	34	39.5	388	2	S36500	E2 protein - human	484	34	39.5	1054	2	A30239	hydroxymethylgluta
412	34	39.5	388	2	AI0208	oligogalacturonide	485	34	39.5	1068	1	A43322	1-phosphatidylinos
413	34	39.5	394	2	S15208	methane monooxygen	486	34	39.5	1068	1	T38110	1-phosphatidylinos
414	34	39.5	399	2	A33396	beta-N-acetylgluco	487	34	39.5	1107	2	S67381	tubulin-folding co
415	34	39.5	402	2	T13614	N-acetyltransferas	488	34	39.5	1204	2	B81947	probable exodeoxyr
416	34	39.5	403	2	D75333	conserved hypotet	489	34	39.5	1204	2	F81158	exodeoxyribonuclea
417	34	39.5	405	2	C75567	adenylosuccinate s	490	34	39.5	1206	2	E86445	hypothetical prote
418	34	39.5	407	2	E88968	protein T27B7.4 [1	491	34	39.5	1239	2	T13809	probable disintegr
419	34	39.5	408	2	AE0103	probable regulator	492	34	39.5	1327	2	D70759	probable otsB prot
420	34	39.5	412	2	AF2979	acyl-CoA dehydroge	493	34	39.5	1369	2	T43433	alpha-glucan synth
421	34	39.5	413	2	T43170	probable triacylgl	494	34	39.5	1385	2	T14158	neurexin IV - mous
422	34	39.5	417	2	F90916	probable transport	495	34	39.5	1474	1	MAHU	alpha-2-macroglobu
423	34	39.5	417	2	F64915	membrane protein y	496	34	39.5	1477	2	T18534	protein-tyrosine k
424	34	39.5	417	2	C85765	probable transport	497	34	39.5	1558	2	C89114	protein C37C3.6a [
425	34	39.5	427	2	A49518	kallistatin precur	498	34	39.5	1573	2	T21219	hypothetical prote
426	34	39.5	432	2	E83357	probable oxidoredu	499	34	39.5	1615	2	B49502	protein-tyrosine-p
427	34	39.5	432	2	S51474	hypothetical prote	500	34	39.5	1710	2	T14005	phospholipase D (E
428	34	39.5	436	2	AI1015	probable exported	501	34	39.5	1767	2	A49502	protein-tyrosine-p
429	34	39.5	443	2	T39540	triglyceride lipas	502	34	39.5	2167	2	T34395	hypothetical prote
430	34	39.5	448	2	F95348	nitric oxide reduc	503	34	39.5	2406	2	A54148	odz protein - frui
431	34	39.5	448	2	AH3095	nitric oxide reduc	504	34	39.5	2469	2	H36812	hypothetical prote
432	34	39.5	448	2	B98191	nitric oxide reduc	505	34	39.5	2471	2	T42977	large tegument pro
433	34	39.5	449	2	AE3634	nitric-oxide reduc	506	34	39.5	2515	2	S47008	tenascin-like prot
434	34	39.5	451	2	JE0166	nitric-oxide reduc	507	34	39.0	180	2	E69269	molybdopterin oxid
435	34	39.5	451	2	G89303	hypothetical prote	508	34	39.0	192	2	T09442	vif protein - huma
436	34	39.5	463	2	D87012	probable glycoly-tr	509	34	39.0	240	1	JQ0807	hydrogenase (EC 1.
437	34	39.5	463	2	T46165	pectate lyase-like	510	34	39.0	240	1	S53656	hydrogenase (EC 1.
438	34	39.5	471	2	T41318	probable glycoosyl	511	34	39.0	255	2	B70982	probable nei prote
439	34	39.5	476	2	F81340	glutamate-ammonia	512	34	39.0	322	1	W2MLE	E2 protein - human
440	34	39.5	481	2	B71929	glutamine syntheta	513	34	39.0	340	2	B88939	protein C05E4.11 [
441	34	39.5	481	2	H64583	glutamine syntheta	514	34	39.0	391	2	S39816	lysine acetyltrans
442	34	39.5	489	2	S62474	probable transcrip	515	34	39.0	567	2	JC5538	Rab geranylgeranyl
443	34	39.5	498	2	H85040	hypothetical prote	516	34	39.0	608	2	T25018	hypothetical prote
444	34	39.5	508	2	T36945	hypothetical prote	517	34	39.0	701	2	F90038	hypothetical prote
445	34	39.5	517	2	B71260	hypothetical prote	518	34	39.0	873	2	JC7079	homeobox protein Z
446	34	39.5	520	2	F70350	recombination prot	519	34	39.0	1094	2	E70697	probable arabinosy
447	34	39.5	533	1	YRMSC5	monophenol monooxy	520	33	38.4	63	2	AI2413	hypothetical prote
448	34	39.5	534	2	S60205	phosphonoacetaldeh	521	33	38.4	69	2	AC3565	hypothetical prote
449	34	39.5	544	2	S52081	diphosphate-fructo	522	33	38.4	75	2	I46471	alpha-actin - rabb
450	34	39.5	546	2	T23920	hypothetical prote	523	33	38.4	83	2	S77417	photochlorophyllid
451	34	39.5	547	2	A32803	glucan 1,4-alpha-m	524	33	38.4	87	2	S71587	ADP-ribosylation f
452	34	39.5	551	2	S05667	glucan 1,4-alpha-m	525	33	38.4	98	2	A39437	expopolysaccharide
453	34	39.5	555	2	D70102	pyrophosphate-fruc	526	33	38.4	98	2	F95975	posttranscription
454	34	39.5	564	2	AF2351	serine/threonine k	527	33	38.4	112	2	T44708	hypothetical prote
455	34	39.5	565	2	AI0479	probable membrane	528	33	38.4	112	2	T44903	hypothetical prote
456	34	39.5	566	2	S15387	malate synthase (E	529	33	38.4	120	2	E49590	Ig heavy chain V r
457	34	39.5	567	1	SYKNMU	malate synthase (E	530	33	38.4	124	2	C97067	hypothetical prote
458	34	39.5	568	1	SYKWA	malate synthase (E	531	33	38.4	129	2	B72332	transposase - Ther
459	34	39.5	573	2	C71312	probable pyrophosp	532	33	38.4	129	2	D72205	transposase - Ther
460	34	39.5	585	2	A46209	protein-tyrosine-p	533	33	38.4	140	2	S03109	actin - pin mould
461	34	39.5	593	1	JN0805	protein-tyrosine-p	534	33	38.4	149	2	T30925	hypothetical prote
462	34	39.5	595	1	A55651	protein-tyrosine-p	535	33	38.4	149	2	G91123	evolved beta-D-gal
463	34	39.5	597	1	A53593	protein-tyrosine-p	536	33	38.4	149	2	F85568	evolved beta-D-gal
464	34	39.5	599	2	A48863	limonene cyclase -	537	33	38.4	149	2	B65096	beta-galactosidase
465	34	39.5	622	2	G90250	glucan 1,4 alpha g	538	33	38.4	153	2	E87313	conserved hypotet
466	34	39.5	631	2	F81227	glucose inhibited	539	33	38.4	157	2	F71906	hypothetical prote
467	34	39.5	641	2	A42019	tyrosine-tRNA liga	540	33	38.4	158	2	I49465	alpha-cardiac acti

541	33	38.4	158	2	P64607	hypothetical prote	614	33	38.4	274	2	AG1678	D-alanyl-D-alanine
542	33	38.4	160	2	S43605	R07E5.13 protein (	615	33	38.4	278	2	C83280	probable transcrip
543	33	38.4	160	2	T16043	hypothetical prote	616	33	38.4	283	2	G89212	conserved hypothet
544	33	38.4	161	2	AG1073	conserved hypothet	617	33	38.4	283	2	E95416	restriction endonu
545	33	38.4	166	2	T37607	probable histone a	618	33	38.4	293	2	S50051	hypothetical prote
546	33	38.4	166	2	AB1150	transcription regu	619	33	38.4	296	2	T77768	hypothetical prote
547	33	38.4	166	2	AC1509	weakly transcripti	620	33	38.4	297	2	S76306	hypothetical prote
548	33	38.4	175	2	B53859	ADP-ribosylation f	621	33	38.4	300	2	G91288	hypothetical prote
549	33	38.4	175	2	E23741	ADP-ribosylation f	622	33	38.4	300	2	B86130	hypothetical prote
550	33	38.4	175	2	T31519	ADP-ribosylation f	623	33	38.4	300	2	S56545	fimbrial protein f
551	33	38.4	175	2	JC4950	ADP-ribosylation f	624	33	38.4	302	2	A32801	fimbrial adhesin p
552	33	38.4	175	2	I50632	hypothetical CPS1	625	33	38.4	305	2	T20685	hypothetical prote
553	33	38.4	176	2	AB0403	probable cytochrom	626	33	38.4	306	2	F83348	probable transcrip
554	33	38.4	180	1	S37599	ADP-ribosylation f	627	33	38.4	308	2	A03000	actin 3 - fruit fi
555	33	38.4	180	2	A53859	ADP-ribosylation f	628	33	38.4	308	2	AE2359	hypothetical prote
556	33	38.4	180	2	S57944	ADP-ribosylation f	629	33	38.4	312	2	S51440	hypothetical prote
557	33	38.4	180	2	A23741	ADP-ribosylation f	630	33	38.4	319	2	H69102	tyrosine-TRNA liga
558	33	38.4	180	2	T32978	ADP-ribosylation f	631	33	38.4	325	2	F94866	probable peroxidase
559	33	38.4	180	2	JC4949	ADP-ribosylation f	632	33	38.4	326	2	PS0017	Ig gamma-1 chain C
560	33	38.4	180	2	JC4948	ADP-ribosylation f	633	33	38.4	326	2	T23139	hypothetical prote
561	33	38.4	180	2	I55371	ADP-ribosylation f	634	33	38.4	327	2	S11452	actin (clone 302)
562	33	38.4	181	2	A41570	ADP-ribosylation f	635	33	38.4	327	2	T30072	hypothetical prote
563	33	38.4	181	2	A33283	ADP-ribosylation f	636	33	38.4	328	2	S05430	actin beta - grass
564	33	38.4	181	2	A45422	ADP-ribosylation f	637	33	38.4	330	2	H64077	aspartate-ammonia
565	33	38.4	181	2	A36167	ADP-ribosylation f	638	33	38.4	331	2	D40649	D-2-hydroxy-acid d
566	33	38.4	181	2	T52339	ADP-ribosylation f	639	33	38.4	331	2	S24409	actin - brown alga
567	33	38.4	181	2	T52341	ADP-ribosylation f	640	33	38.4	332	2	S44206	hypothetical prote
568	33	38.4	181	2	S49325	ADP-ribosylation f	641	33	38.4	333	2	S15238	O-antigen acetatylas
569	33	38.4	181	2	T15341	ADP-ribosylation f	642	33	38.4	333	2	T52594	squamosa promoter
570	33	38.4	181	2	S66337	ADP-ribosylation f	643	33	38.4	336	2	T04085	actin - maize (fra
571	33	38.4	181	2	JC4946	ADP-ribosylation f	644	33	38.4	339	2	T21473	hypothetical prote
572	33	38.4	181	2	JC4947	ADP-ribosylation f	645	33	38.4	347	2	A61988	probable N-acetyl-
573	33	38.4	181	2	JC4945	ADP-ribosylation f	646	33	38.4	347	2	D81043	N-acetyl-gamma-glu
574	33	38.4	181	2	T48021	ADP-ribosylation f	647	33	38.4	347	2	T37995	probable fatty aci
575	33	38.4	181	2	G96728	probable ADP-ribos	648	33	38.4	348	1	BVECMCB	mcrC protein - Esc
576	33	38.4	181	2	S28875	ADP-ribosylation f	649	33	38.4	349	2	B25819	actin, fetal skele
577	33	38.4	182	2	A49520	ADP-ribosylation f	650	33	38.4	350	2	A54420	beta-galactoside a
578	33	38.4	182	2	C49993	ADP-ribosylation f	651	33	38.4	351	1	G69290	probable hexosyltr
579	33	38.4	183	2	D49993	ADP-ribosylation f	652	33	38.4	361	2	S68089	actin 2 - Arabidop
580	33	38.4	186	2	E69537	conserved hypothet	653	33	38.4	362	2	A26559	actin type 5, cyto
581	33	38.4	188	2	E86368	F508 5 protein - A	654	33	38.4	362	2	S68090	actin 8 - Arabidop
582	33	38.4	188	2	T48640	ADP-ribosylation f	655	33	38.4	365	2	A37431	actin, type 1 - Em
583	33	38.4	190	2	H83172	hypothetical prote	656	33	38.4	365	2	S49007	actin - Pyrhium ir
584	33	38.4	190	2	A82343	conserved hypothet	657	33	38.4	367	2	JT0596	actin Ardd - slime
585	33	38.4	192	1	C44001	vif protein - huma	658	33	38.4	370	2	A29664	actin beta - bovin
586	33	38.4	192	2	S42988	viral infectivity	659	33	38.4	374	1	ATBOB	actin gamma - bovi
587	33	38.4	192	2	T39367	hypothetical prote	660	33	38.4	374	1	ATBOG	gamma-actin - huma
588	33	38.4	193	2	AG3371	hypothetical prote	661	33	38.4	374	2	JC5818	pectate lyase (EC
589	33	38.4	194	2	A70838	hypothetical prote	662	33	38.4	375	1	WZWC6C	actin, aortic smoo
590	33	38.4	195	2	S39777	actin beta - pig (	663	33	38.4	375	1	ATBOSM	actin, skeletal mu
591	33	38.4	195	2	S20097	actin 85c - potato	664	33	38.4	375	1	ATRB	actin beta - rat
592	33	38.4	197	2	S36453	ADP-ribosylation f	665	33	38.4	375	1	ATRTC	actin beta - cytosk
593	33	38.4	201	2	AI1106	conserved hypothet	666	33	38.4	375	1	A48324	actin - Acanthamo
594	33	38.4	201	2	AE1468	hypothetical prote	667	33	38.4	375	1	ATAX	actin - yeast (Sac
595	33	38.4	204	2	T04658	hypothetical prote	668	33	38.4	375	1	ATBY	actin beta - chick
596	33	38.4	205	2	T07423	actin - Chlorella	669	33	38.4	375	1	ATCHB	actin - slime mold
597	33	38.4	213	2	A61043	actin CA15 - sea s	670	33	38.4	375	1	ATDO	actin beta - human
598	33	38.4	213	2	D69938	hemolysin III homo	671	33	38.4	375	1	ATHUB	actin gamma 1 - hu
599	33	38.4	225	2	H75571	conserved hypothet	672	33	38.4	375	1	ATHUG	actin gamma 1 - hu
600	33	38.4	225	2	S73585	MG068 homolog D02	673	33	38.4	375	1	ATMSB	actin beta - mouse
601	33	38.4	231	2	H86463	Fl2G12.17 protein	674	33	38.4	375	1	ATMSG	actin gamma - mous
602	33	38.4	232	2	A42095	floral homeotic pr	675	33	38.4	375	1	ATRB	actin beta, non-mu
603	33	38.4	233	2	S63660	hypothetical prote	676	33	38.4	375	1	ATZM1	actin - maize
604	33	38.4	241	2	D71167	hypothetical prote	677	33	38.4	375	1	JS0702	actin - yeast (Sac
605	33	38.4	247	2	G82202	hypothetical prote	678	33	38.4	375	1	S11222	actin gamma, cytos
606	33	38.4	252	2	AF1861	ATP-binding protei	679	33	38.4	375	2	A26836	actin - yeast (Klu
607	33	38.4	258	2	E90126	hypothetical prote	680	33	38.4	375	2	A55001	actin - fission ye
608	33	38.4	260	2	I51544	MHC class II beta-	681	33	38.4	375	2	A54728	actin beta - goose
609	33	38.4	261	2	E72735	hypothetical prote	682	33	38.4	375	2	S33386	actin alpha, cardi
610	33	38.4	273	2	B84868	probable endochiti	683	33	38.4	375	2	JS0385	actin, cytosolic (
611	33	38.4	273	2	B64608	outer membrane pro	684	33	38.4	375	2	JS0385	actin gamma - Emer
612	33	38.4	274	2	C71300	hypothetical prote	685	33	38.4	375	2	JS0385	actin 1 - Pneumocy
613	33	38.4	274	2	AG1306	D-alanyl-D-alanine	686	33	38.4	375	2	T25272	hypothetical prote



687	33	38.4	375	2	S71125	actin beta-2, cyto	760	33	38.4	377	1	ATHU	actin alpha 1, ske
688	33	38.4	375	2	S71124	actin beta-1, cyto	761	33	38.4	377	1	ATHUC	actin, cardiac mus
689	33	38.4	375	2	S71126	actin beta, cyto	762	33	38.4	377	1	ATHUSM	actin alpha 2, aor
690	33	38.4	375	2	S42103	actin - Puccinia g	763	33	38.4	377	1	ATMUM1	actin - Arabidopsi
691	33	38.4	375	2	S03126	actin - imperfect	764	33	38.4	377	1	ATRBMS	actin alpha, smoot
692	33	38.4	375	2	S70377	actin - Phaffia rh	765	33	38.4	377	1	ATRT	actin, skeletal mu
693	33	38.4	375	2	E82681	conserved hypothet	766	33	38.4	377	1	ATRZ1	actin 1 - rice
694	33	38.4	376	1	ATFF87	actin 7 - fruit fl	767	33	38.4	377	1	B29686	actin alpha, cardi
695	33	38.4	376	1	ATSY3	actin - soybean	768	33	38.4	377	2	B24848	actin alpha-3, ske
696	33	38.4	376	1	A43552	actin gamma, cytos	769	33	38.4	377	2	B24848	actin alpha-1, car
697	33	38.4	376	1	ATAXE	actin - Entamoeba	770	33	38.4	377	2	A29686	actin alpha-2, ske
698	33	38.4	376	1	ATCHSM	actin gamma, smoot	771	33	38.4	377	2	JCS301	skeletal alpha-act
699	33	38.4	376	1	ATFF8	actin 8 - fruit fl	772	33	38.4	377	2	S31933	actin - common tob
700	33	38.4	376	1	ATFY	actin - slime mold	773	33	38.4	377	2	S20093	actin 101 - potato
701	33	38.4	376	1	ATRY3	actin 3 - rice	774	33	38.4	377	2	S20098	actin 97 - potato
702	33	38.4	376	1	ATRZ7	actin 7 - rice	775	33	38.4	377	2	S20096	actin 75 - potato
703	33	38.4	376	1	ATURS	actin Cyl - sea ur	776	33	38.4	377	2	S20094	actin 58 - potato
704	33	38.4	376	1	E69957	gamma-D-glutamyl-L	777	33	38.4	377	2	S20095	actin 71 - potato
705	33	38.4	376	2	B23412	actin 12 - slime m	778	33	38.4	377	2	JE0147	actin 1 - sorghum
706	33	38.4	376	2	A48449	Actin-1A - nematod	779	33	38.4	377	2	S71120	actin alpha, cardi
707	33	38.4	376	2	A27724	actin 1 - Trypanos	780	33	38.4	377	2	S71118	actin alpha-1, ske
708	33	38.4	376	2	B27724	actin 2 - Trypanos	781	33	38.4	377	2	S71119	actin alpha-2, ske
709	33	38.4	376	2	A45634	actin - Cryptospor	782	33	38.4	377	2	S68107	actin 7 - Arabidop
710	33	38.4	376	2	A54596	actin I - malaria	783	33	38.4	377	2	S68112	actin 3 [imported]
711	33	38.4	376	2	A54509	actin II - malaria	784	33	38.4	377	2	S68110	actin 12 - Arabido
712	33	38.4	376	2	A29407	actin - Tetrahymen	785	33	38.4	377	2	S68108	actin 4 - Arabido
713	33	38.4	376	2	A40261	actin gamma, enter	786	33	38.4	377	2	S68109	actin 11 - Arabido
714	33	38.4	376	2	C23412	actin 3-sub1 - sli	787	33	38.4	377	2	S58316	actin - garden pea
715	33	38.4	376	2	A25084	actin 15 - slime m	788	33	38.4	377	2	T51177	actin [imported] -
716	33	38.4	376	2	A44940	actin - pork tapew	789	33	38.4	377	2	T51178	actin AC12 [import
717	33	38.4	376	2	A32788	actin gamma, smoot	790	33	38.4	377	2	T51181	actin 1 [imported]
718	33	38.4	376	2	A31375	actin, smooth musc	791	33	38.4	377	2	T51183	actin isoform B [i
719	33	38.4	376	2	A25135	actin A3, cytosoli	792	33	38.4	377	2	T51180	actin [imported] -
720	33	38.4	376	2	S04538	actin 878 - fruit	793	33	38.4	377	2	T51175	actin [imported] -
721	33	38.4	376	2	JC1246	actin - fruit fly	794	33	38.4	377	2	T51184	actin [imported] -
722	33	38.4	376	2	S12628	actin - malaria pa	795	33	38.4	377	2	T51179	actin [imported] -
723	33	38.4	376	2	S07284	actin - Tetrahymen	796	33	38.4	377	2	T51176	actin [imported] -
724	33	38.4	376	2	JS0190	actin, cytosolic -	797	33	38.4	377	2	T51182	actin [imported] -
725	33	38.4	376	2	S07288	actin, muscle - st	798	33	38.4	378	1	ATSY1	actin 1 - soybean
726	33	38.4	376	2	S09578	actin 15A - sea ur	799	33	38.4	378	2	H84849	probable actin [im
727	33	38.4	376	2	S09578	actin - sea urchin	800	33	38.4	379	1	ATRZ2	actin 2 - rice
728	33	38.4	376	2	JQ0154	actin - Hydra atte	801	33	38.4	379	2	S33387	actin, muscle - se
729	33	38.4	376	2	JN0833	actin (clone gen3)	802	33	38.4	380	2	D23412	actin 3-sub2 - sli
730	33	38.4	376	2	JN0832	actin (clones Ia a	803	33	38.4	380	2	S07002	actin 1 - carrot
731	33	38.4	376	2	S24408	actin - Achlya bis	804	33	38.4	380	2	A83458	class I histocampa
732	33	38.4	376	2	JE0414	actin A - Phytoph	805	33	38.4	381	2	S35940	bacteroid developm
733	33	38.4	376	2	S27135	actin 4 - Caenorha	806	33	38.4	382	2	AG2859	hypothetical prote
734	33	38.4	376	2	S16710	actin 1 and actin	807	33	38.4	383	2	A12748	probable ATP-bindi
735	33	38.4	376	2	S16709	actin 2 - Caenorha	808	33	38.4	383	2	H97529	hypothetical prote
736	33	38.4	376	2	T24448	hypothetical prote	809	33	38.4	386	2	T06788	actin - garden pea
737	33	38.4	376	2	JCS228	actin 2 - earthwor	810	33	38.4	387	2	D88968	protein T27B7.3 li
738	33	38.4	376	2	JCS227	actin 1 - earthwor	811	33	38.4	389	2	T32516	hypothetical prote
739	33	38.4	376	2	S12730	actin - California	812	33	38.4	394	2	A81789	probable monooxyge
740	33	38.4	376	2	S43509	actin - California	813	33	38.4	394	2	F81212	UbiH family protei
741	33	38.4	376	2	S11450	actin (clone 205)	814	33	38.4	395	2	A13402	salicylate 1-mono
742	33	38.4	376	2	S11451	actin (clone 211)	815	33	38.4	400	2	S07733	NADH2 dehydrogenas
743	33	38.4	376	2	S11453	actin (clone 403)	816	33	38.4	402	1	RERTK	renin (EC 3.4.23.1
744	33	38.4	376	2	S49481	actin 5 - Atlantic	817	33	38.4	403	2	A72523	probable acyl-CoA
745	33	38.4	376	2	S49479	actin 11 - Atlanti	818	33	38.4	413	2	H86823	ammonium transport
746	33	38.4	376	2	S49480	actin 3 - Atlantic	819	33	38.4	419	2	E83904	hypothetical prote
747	33	38.4	376	2	S71123	actin alpha-anomal	820	33	38.4	422	2	E97636	bacteroid developm
748	33	38.4	376	2	S01077	actin beta, cytosk	821	33	38.4	423	2	A64486	dihydroorotase (EC
749	33	38.4	376	2	S25488	actin 1 - garden p	822	33	38.4	427	2	A64329	phosphopyruvate hy
750	33	38.4	376	2	S26435	actin 2 - garden p	823	33	38.4	427	2	T45915	actin (ACT3) - Ara
751	33	38.4	376	2	S07639	actin - Yeast (Can	824	33	38.4	429	2	T14237	hypothetical prote
752	33	38.4	376	2	JCS750	actin A4 - silkwor	825	33	38.4	431	2	T41560	phenylalanyl-trna
753	33	38.4	376	2	S09059	actin A1 - silkwor	826	33	38.4	434	2	S77330	hypothetical prote
754	33	38.4	376	2	S07382	actin A2 - silkwor	827	33	38.4	434	2	G70011	conserved hypothet
755	33	38.4	377	1	A25719	actin alpha, aorti	828	33	38.4	435	2	G87334	acyl-CoA dehydroge
756	33	38.4	377	1	A22224	actin alpha, vascu	829	33	38.4	436	2	G85749	aminoacylase (EC 3
757	33	38.4	377	1	A23022	actin, cardiac mus	830	33	38.4	441	2	T35169	probable glucosida
758	33	38.4	377	1	A24904	actin alpha, skele	831	33	38.4	441	2	E64883	probable aminohydr
759	33	38.4	377	1	ATCH	actin alpha, skele	832	33	38.4	441	2	B90869	probable aminohydr

833	33	38.4	443	2	T21598	hypotheical prote	906	33	38.4	761	2	A10368	malate dehydrogena
834	33	38.4	447	2	C85085	hypotheical prote	907	33	38.4	762	2	JC7174	N,N-dimethylformam
835	33	38.4	447	2	AD2474	hypotheical prote	908	33	38.4	765	2	JC7498	hypotheical prote
836	33	38.4	454	1	T25203	probable membrane-	909	33	38.4	794	2	T39171	probable peroxisom
837	33	38.4	462	2	G01804	interleukin 3-regu	910	33	38.4	800	2	T26683	hypotheical prote
838	33	38.4	463	2	T09243	dnak-type molecula	911	33	38.4	816	2	H85028	probable GTP pyrop
839	33	38.4	463	2	D87624	sodium-galactoside	912	33	38.4	828	1	D39142	outer membrane ush
840	33	38.4	473	1	WMBE51	Uti10 protein - hum	913	33	38.4	837	1	S54429	outer membrane ush
841	33	38.4	484	2	AD0431	xylokinaase (EC 2	914	33	38.4	855	2	JC7731	membrane-bound arg
842	33	38.4	493	1	ACMSE	nicotinic acetylch	915	33	38.4	859	1	S65938	nitrate reductase
843	33	38.4	493	2	T34453	hypotheical prote	916	33	38.4	862	2	S64821	probable membrane
844	33	38.4	495	2	S71900	RNA-directed DNA p	917	33	38.4	871	2	G84601	protein kinase k
845	33	38.4	501	2	S16711	ABC1 protein precu	918	33	38.4	878	2	T08559	protein kinase hom
846	33	38.4	505	2	T22258	hypotheical prote	919	33	38.4	937	2	G87640	TonB-dependent rec
847	33	38.4	506	2	T07942	probable squalene	920	33	38.4	937	2	A56517	nucleoporin Nup98
848	33	38.4	511	2	T15629	hypotheical prote	921	33	38.4	938	2	S20480	traG protein - Esc
849	33	38.4	514	1	ALBSN	alpha-amylase (EC	922	33	38.4	955	2	T39765	probable nuclear m
850	33	38.4	515	2	JC5458	inulinase (EC 3.2.	923	33	38.4	1009	2	C64483	hypotheical prote
851	33	38.4	516	2	D64410	replication factor	924	33	38.4	1060	1	P2XRA4	outer capsid prote
852	33	38.4	516	2	H72427	DNA mismatch repai	925	33	38.4	1090	2	D72048	pbp2-transglycolas
853	33	38.4	516	2	JE0301	inulinase (EC 3.2.	926	33	38.4	1113	2	D86142	hypotheical prote
854	33	38.4	529	1	YRHU1	monophenol monooxy	927	33	38.4	1186	2	AG1928	two-component hybr
855	33	38.4	529	2	B49993	glycylpeptide N-te	928	33	38.4	1217	2	F97177	alpha-glucosidase
856	33	38.4	530	2	AC2085	phosphodiesterase/	929	33	38.4	1232	2	T43027	neural cell adhesi
857	33	38.4	532	2	JC1392	monophenol monooxy	930	33	38.4	1271	2	D64237	hypotheical prote
858	33	38.4	533	2	T32389	hypotheical prote	931	33	38.4	1289	1	GUBPT4	proximal tail fibe
859	33	38.4	535	2	G95155	hypotheical prote	932	33	38.4	1313	2	T29027	hypotheical prote
860	33	38.4	535	2	B98022	hypotheical prote	933	33	38.4	1397	2	T51292	Dna2p - fission ye
861	33	38.4	537	2	H88087	protein B0454.4 [i	934	33	38.4	1398	2	T39568	hypotheical helic
862	33	38.4	539	2	S53529	monophenol monooxy	935	33	38.4	1449	2	T30857	glucosyltransferas
863	33	38.4	545	2	T02279	hypotheical prote	936	33	38.4	1449	2	T30552	glucosyltransferas
864	33	38.4	555	1	S17502	inulinase (EC 3.2.	937	33	38.4	1518	2	A44811	glucosyltransferas
865	33	38.4	556	1	S31330	inulinase (EC 3.2.	938	33	38.4	1562	2	T17411	polyketide synthas
866	33	38.4	562	2	S16594	regulatory protein	939	33	38.4	1790	1	S27772	vitellogenin precu
867	33	38.4	567	2	D84400	phenylalanyl-tRNA	940	33	38.4	1928	2	JS0610	beta-galactosidase
868	33	38.4	584	2	H86531	C7085 hypotheical	941	33	38.4	2470	2	I50726	cation-independent
869	33	38.4	584	2	I50419	s-glicerol precursor	942	33	38.4	2817	2	B37033	uncharacterized pr
870	33	38.4	584	2	A72092	conserved hypothet	943	33	38.4	2958	2	S64921	probable membrane
871	33	38.4	586	2	S65802	transcription acti	944	33	38.4	3034	2	T14119	seven-pass transme
872	33	38.4	586	2	T15259	hypotheical prote	945	33	38.4	3305	2	T18358	apolipophorin prec
873	33	38.4	598	2	S51456	probable membrane	946	33	38.4	133	2	B84087	hypotheical prote
874	33	38.4	598	2	S66954	probable membrane	947	33	38.4	183	1	I64006	hypotheical prote
875	33	38.4	599	2	JN0818	transferrin-bindin	948	33	38.4	192	2	S42944	viral infectivity
876	33	38.4	599	2	S67084	probable membrane	949	33	38.4	192	2	S43003	viral infectivity
877	33	38.4	605	2	G85651	probable membrane	950	33	38.4	192	2	S42950	conserved hypothet
878	33	38.4	605	2	D90791	probable membrane	951	33	38.4	192	2	AB0726	conserved hypothet
879	33	38.4	610	2	A41388	Lc regulatory prot	952	33	38.4	199	2	AC3120	superoxide dismuta
880	33	38.4	612	2	S60702	regulatory protein	953	33	38.4	200	1	DSHSNH	superoxide dismuta
881	33	38.4	615	2	T01422	GTP pyrophosphokin	954	33	38.4	200	2	B84288	superoxide dismuta
882	33	38.4	616	2	S20089	gene SN protein -	955	33	38.4	202	2	T50047	superoxide dismuta
883	33	38.4	620	2	S52494	protein kinase hom	956	33	38.4	202	2	A83826	superoxide dismuta
884	33	38.4	627	1	S27393	sphingomyelin phos	957	33	38.4	211	2	B75415	superoxide dismuta
885	33	38.4	632	2	AE3560	proline/betaine tr	958	33	38.4	226	2	B69709	superoxide dismuta
886	33	38.4	634	2	T01408	inclusion protein	959	33	38.4	240	2	D98167	superoxide dismuta
887	33	38.4	649	2	G71530	probable ATP synth	960	33	38.4	261	2	T11512	cytochrome-c oxida
888	33	38.4	649	2	C81687	ATP synthase, chai	961	33	38.4	296	2	JC7283	hydroxyarylamine s
889	33	38.4	651	2	D72605	probable long-chai	962	33	38.4	296	2	JC7282	xyloglucan endo-1,
890	33	38.4	659	2	AB0731	probable hydrolase	963	33	38.4	296	2	D49539	xyloglucan endo-1,
891	33	38.4	660	2	E81549	ATP synthase, chai	964	33	38.4	339	1	NCEPX4	exonuclease 47 (EC
892	33	38.4	660	2	C86502	ATP synthase subun	965	33	38.4	365	2	AF3217	DNA-damage-inducib
893	33	38.4	660	2	E72121	ATP synthase chain	966	33	38.4	412	2	AG0177	probable hydroxyme
894	33	38.4	662	2	S50319	hypotheical F-box	967	33	38.4	428	2	B82248	long-chain fatty a
895	33	38.4	663	1	TMVVR	protein-tyrosine k	968	33	38.4	449	2	F83627	hypotheical prote
896	33	38.4	671	2	S51599	Om(2D) protein - f	969	33	38.4	555	2	AE2147	cytochrome c oxida
897	33	38.4	694	2	S71786	wingless receptor	970	33	38.4	648	2	S10869	enterotoxin A - Cl
898	33	38.4	704	2	S46000	probable membrane	971	33	38.4	682	2	JQ0420	beta-1,3-glucanase
899	33	38.4	707	2	T09340	hypotheical prote	972	33	38.4	787	1	JDVLW2	DNA-directed DNA p
900	33	38.4	721	2	H82528	L-ascorbate oxidas	973	33	38.4	859	2	AC2089	adenylate cyclase
901	33	38.4	717	2	S47857	basic protein, cyt	974	33	38.4	1624	2	C70867	probable Helix-tur
902	33	38.4	746	2	S71892	RNA-directed DNA p	975	33	38.4	19	2	S43657	hsp90 protein homo
903	33	38.4	751	1	S26071	photosystem I prot	976	33	38.4	68	2	AG2291	hypotheical prote
904	33	38.4	757	2	JS0198	genome polyprotein	977	33	38.4	70	2	T04408	actin - barley (fr
905	33	38.4	760	2	A99233	hypotheical prote	978	33	38.4	81	2	A24522	mitochondrial prot

979 32 37.2 85 2 B47354 ysdF protein - Sal  
980 32 37.2 86 2 G82801 hypothetical prote  
981 32 37.2 89 2 D64465 hypothetical prote  
982 32 37.2 91 2 I48658 dm1 antigen - mous  
983 32 37.2 94 2 T29829 hypothetical prote  
984 32 37.2 101 2 B38189 protein-tyrosine-p  
985 32 37.2 107 2 AC3315 hypothetical membr  
986 32 37.2 108 2 B30352 protein-lysine 6-o  
987 32 37.2 109 2 PC2185 heat shock protein  
988 32 37.2 111 2 F69779 transcription regu  
989 32 37.2 117 2 F83449 conserved hypotet  
990 32 37.2 119 2 S53764 hypothetical prote  
991 32 37.2 120 2 AH0858 probable 6-pyruvoy  
992 32 37.2 122 2 G82217 probable 6-pyruvoy  
993 32 37.2 125 2 B83857 chorismate mutase  
994 32 37.2 129 2 S76606 hypothetical prote  
995 32 37.2 130 2 AI2467 50S ribosomal prot  
996 32 37.2 135 2 T40220 hypothetical prote  
997 32 37.2 136 2 B86628 prophage p81 prote  
998 32 37.2 136 2 AI0124 probable prepin  
999 32 37.2 140 2 S54084 probable membrane  
1000 32 37.2 140 2 C95999 hypothetical prote

## ALIGNMENTS

RESULT 1  
F71190  
A;Title: probable chromosome assembly protein - Pyrococcus horikoshii  
C;Species: Pyrococcus horikoshii  
C;Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 12-Jul-2004  
C;Accession: F71190  
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a  
A;Reference number: A71000; MUID:98344137; PMID:9679194  
A;Accession: F71190  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1179 <KAW>  
A;Cross-references: UNIPROT:O59462; UNIPARC:UPI000006688D; GB:AP000007; NID:g2326134; PT  
A;Experimental source: strain OT3  
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C;Genetics:  
A;Gene: PH1798

Query Match 59.3%; Score 51; DB 2; Length 1179;  
Best Local Similarity 81.8%; Pred. No. 2.5;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 HYLLNGRTATR 14  
||||:|||||  
Db 113 HYWLNGRRATR 123

RESULT 2  
JC5251  
A;Title: beta-galactoside alpha-2,3-sialyltransferase (EC 2.4.99.4) - human  
C;Species: Homo sapiens (man)  
C;Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 05-Oct-2004  
C;Accession: JC5251; G01021  
R;Kim, Y.J.; Kim, K.S.; Kim, S.H.; Kim, C.H.; Ko, J.H.; Choe, I.S.; Tsuji, S.; Lee, Y.C.  
Biochem. Biophys. Res. Commun. 228, 324-327, 1996  
A;Title: Molecular cloning and expression of human Gal beta 1,3GalNAc alpha 2,3-sialyltra  
A;Reference number: JC5251; MUID:97079181; PMID:8920913  
A;Accession: JC5251  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-350 <KIM>  
A;Cross-references: UNIPROT:Q16842; UNIPARC:UPI0000001C5C; GB:U63090; NID:g1773282; PIDN  
A;Experimental source: liver

R;Giordanengo, V.  
submitted to the EMBL Data Library, March 1996  
A;Reference number: H00561  
A;Accession: G01021  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 1-350 <GIO>  
A;Cross-references: UNIPARC:UPI0000001C5C; EMBL:X96667; NID:g1235530  
C;Comment: This enzyme catalyzes the transfer of sialic acid from CMP-NeuAc to the termin  
substrate preference for glycolipid than for O-linked oligosaccharides of glycoproteins.  
C;Genetics:  
A;Gene: ST3(0)-II  
C;Superfamily: sialyltransferase  
C;Keywords: glycosyltransferase

Query Match 53.5%; Score 46; DB 2; Length 350;  
Best Local Similarity 63.6%; Pred. No. 4.7;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHHYLLNGRTA 12  
||||:|||||  
Db 307 WHHYWNNRYA 317

RESULT 3  
B54420  
A;Title: beta-galactoside alpha-2,3-sialyltransferase (EC 2.4.99.4) ST3GALNA.2 - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 05-Oct-2004  
C;Accession: B54420  
J;Lee, Y.C.; Kojima, N.; Wada, E.; Kurosawa, N.; Nakaoka, T.; Hamamoto, T.; Tsuji, S.  
J. Biol. Chem. 269, 10028-10033, 1994  
A;Title: Cloning and expression of cDNA for a new type of Galbeta1,6GalNAc alpha2,3-sialyl  
A;Reference number: A54420; MUID:94193584; PMID:8144500  
A;Accession: B54420  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-350 <LEE>  
A;Cross-references: UNIPROT:Q11205; UNIPARC:UPI000013596B; GB:X76988; NID:g475225; PIDN:  
C;Superfamily: sialyltransferase  
C;Keywords: glycosyltransferase; transmembrane protein

Query Match 53.5%; Score 46; DB 2; Length 350;  
Best Local Similarity 63.6%; Pred. No. 4.7;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHHYLLNGRTA 12  
||||:|||||  
Db 307 WHHYWNNRYA 317

RESULT 4  
T01245  
A;Title: N-acetyltransferase homolog Fl6M14.6 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004  
C;Accession: T01245; C84801  
R;Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.  
submitted to the EMBL Data Library, July 1998  
A;Description: Arabidopsis thaliana chromosome II BAC Fl6M14 genomic sequence.  
A;Reference number: Z14213  
A;Accession: T01245  
A;Status: translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1-190 <ROU>  
A;Cross-references: UNIPROT:O80438; UNIPARC:UPI0000009EFCB; EMBL:AC003028; NID:g3335356; I  
A;Experimental source: cultivar Columbia  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: C84801  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-190 <STO>  
A;Cross-references: UNIPARC:UPI000009EFCB; GB:AF002093; NID:G3335361; PIDN:AAC27162.1; C  
C;Species: Gallus gallus (chicken)  
C;Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 05-Oct-2004  
C;Accession: S55675  
R;Kurosawa, N.; Hamamoto, T.; Inoue, M.; Tsuji, S.  
Biochim. Biophys. Acta 1244, 216-222, 1995  
A;Title: Molecular cloning and expression of chick Gal-beta-1,3GalNAc alpha-2,3-sialyltr  
A;Reference number: S55675; MUID:95284088; PMID:7766661  
A;Accession: S55675  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-342 <KUR>  
A;Cross-references: UNIPROT:Q11200; UNIPARC:UPI0000135969; GB:X80503; NID:g975654; PIDN:  
C;Superfamily: sialyltransferase

Query Match 50.0%; Score 43; DB 2; Length 342;  
Best Local Similarity 54.5%; Pred. No. 14;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHYYLNGRTA 12  
||||| :|

A;Accession: C84801  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-190 <STO>  
A;Cross-references: UNIPARC:UPI000009EFCB; GB:AF002093; NID:G3335361; PIDN:AAC27162.1; C  
C;Species: Nostoc sp. PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 31-Dec-2004  
C;Accession: AC2030  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AC2030  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-290 <KUR>  
A;Cross-references: UNIPROT:Q8YM23; UNIPARC:UPI000000CE1F1; GB:BA000019; PIDN:BAB73492.1;  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: alr1793  
C;Superfamily: nodulation protein nodB

Query Match 51.2%; Score 44; DB 2; Length 290;  
Best Local Similarity 53.8%; Pred. No. 8.2;  
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 TWHHYLLNGRTAT 13  
||||| :|

Db 159 TWHHYRMRMEAT 171  
||||| :|

RESULT 6  
S55675  
Gal-beta-1,3GalNAc alpha-2,3-sialyltransferase - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 05-Oct-2004  
C;Accession: S55675  
R;Kurosawa, N.; Hamamoto, T.; Inoue, M.; Tsuji, S.  
Biochim. Biophys. Acta 1244, 216-222, 1995  
A;Title: Molecular cloning and expression of chick Gal-beta-1,3GalNAc alpha-2,3-sialyltr  
A;Reference number: S55675; MUID:95284088; PMID:7766661  
A;Accession: S55675  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-342 <KUR>  
A;Cross-references: UNIPROT:Q11200; UNIPARC:UPI0000135969; GB:X80503; NID:g975654; PIDN:  
C;Superfamily: sialyltransferase

Query Match 50.0%; Score 43; DB 2; Length 342;  
Best Local Similarity 54.5%; Pred. No. 14;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHYYLNGRTA 12  
||||| :|

A;Accession: C84801  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-190 <STO>  
A;Cross-references: UNIPARC:UPI000009EFCB; GB:AF002093; NID:G3335361; PIDN:AAC27162.1; C  
C;Species: Nostoc sp. PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 31-Dec-2004  
C;Accession: AC2030  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AC2030  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-290 <KUR>  
A;Cross-references: UNIPROT:Q8YM23; UNIPARC:UPI000000CE1F1; GB:BA000019; PIDN:BAB73492.1;  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: alr1793  
C;Superfamily: nodulation protein nodB

Query Match 51.2%; Score 44; DB 2; Length 290;  
Best Local Similarity 53.8%; Pred. No. 8.2;  
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 TWHHYLLNGRTAT 13  
||||| :|

Db 159 TWHHYRMRMEAT 171  
||||| :|

RESULT 6  
S55675  
Gal-beta-1,3GalNAc alpha-2,3-sialyltransferase - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 05-Oct-2004  
C;Accession: S55675  
R;Kurosawa, N.; Hamamoto, T.; Inoue, M.; Tsuji, S.  
Biochim. Biophys. Acta 1244, 216-222, 1995  
A;Title: Molecular cloning and expression of chick Gal-beta-1,3GalNAc alpha-2,3-sialyltr  
A;Reference number: S55675; MUID:95284088; PMID:7766661  
A;Accession: S55675  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-342 <KUR>  
A;Cross-references: UNIPROT:Q11200; UNIPARC:UPI0000135969; GB:X80503; NID:g975654; PIDN:  
C;Superfamily: sialyltransferase

Query Match 50.0%; Score 43; DB 2; Length 342;  
Best Local Similarity 54.5%; Pred. No. 14;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHYYLNGRTA 12  
||||| :|

Db 299 WHYYWENNAS 309

RESULT 7  
D85330  
hypothetical protein AT4g28370 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C;Accession: D85330  
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A;Reference number: A85001; MUID:20083488; PMID:10617198  
A;Accession: D85330  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-479 <STO>  
A;Cross-references: UNIPROT:O49446; UNIPARC:UPI000009FD49; GB:NC\_001368; NID:G7369691; P1  
C;Genetics:  
A;Gene: AT4g28370  
A;Map position: 4

Query Match 50.0%; Score 43; DB 2; Length 479;  
Best Local Similarity 66.7%; Pred. No. 21;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 HHYYLNGRTAT 14  
||||| :|

Db 330 HPYYILGWTAT 341  
||||| :|

RESULT 8  
JC7506  
heparanase protein 2a - human  
C;Species: Homo sapiens (man)  
C;Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 09-Jul-2004  
C;Accession: JC7506  
R;McKenzie, E.; Iysson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hircock, M.; Pat  
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000  
A;Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family me  
A;Reference number: JC7506  
A;Accession: JC7506  
A;Molecule type: mRNA  
A;Residues: 1-480 <MCK>  
A;Cross-references: UNIPROT:Q9HB39; UNIPARC:UPI000003E88A; GB:AF282885  
C;Comment: This protein, a intracellular membrane-bound enzyme, has biological and therai  
therapies.  
C;Genetics:  
A;Gene: hpa2a  
A;Map position: 10q23-10q24  
C;Keywords: heparin binding; membrane bound

Query Match 50.0%; Score 43; DB 2; Length 480;  
Best Local Similarity 60.0%; Pred. No. 21;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 TWHHYLLNGR 10  
||||| :|

Db 220 TWQHCYIDGR 229  
||||| :|

RESULT 9  
T04607  
hypothetical protein F2009.50 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 21-May-1999  
C;Accession: T04607  
R;Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.X.  
submitted to the Protein Sequence Database, October 1998  
A;Reference number: Z15380  
A;Accession: T04607  
A;Molecule type: DNA  
A;Residues: 1-481 <BEV>

A;Cross-references: UNIPARC:UPI0000179E9C; EMBL:AL021749  
A;Experimental source: cultivar Columbia; BAC clone F2009  
C;Genetics:  
A;Map position: 4  
A;Introns: 4/3; 29/1; 82/3; 106/1; 128/2; 177/1; 226/3; 323/3; 420/1; 457/1  
A;Note: F2009.50

Query Match 50.0%; Score 43; DB 2; Length 481;  
Best Local Similarity 66.7%; Pred. No. 21;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 3 HHYILNGRTATR 14  
| | | | |  
| | | | |  
DB 259 HPYILNGMTATR 270

RESULT 10  
B75150  
chromosome segregation protein (smc1) PAB2109 - *Pyrococcus abyssi* (strain Orsay)  
C;Species: *Pyrococcus abyssi*  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004  
C;Accession: B75150  
R;Anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A;Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome stru  
A;Reference number: A75001  
A;Accession: B75150  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1177 <KAW>  
A;Cross-references: UNIPROT:Q9V1R8; UNIPARC:UPI00000346A8; GB:AJ248284; GB:AL096836; NID  
A;Experimental source: strain Orsay  
C;Genetics:  
A;Gene: PAB2109  
C;Superfamily: chromosome segregation protein SMC1

Query Match 50.0%; Score 43; DB 2; Length 1177;  
Best Local Similarity 80.0%; Pred. No. 55;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 5 YYLNGRTATR 14  
| | | | |  
| | | | |  
DB 114 YWNGRRATR 123

RESULT 11  
S77524  
chromosome segregation protein smc1 - *Synechocystis* sp. (strain PCC 6803)  
N;Alternate names: protein sll1120  
C;Species: *Synechocystis* sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C;Accession: S77524  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
s.  
A;Reference number: S74322; MUID:97061201; PMID:8905231  
A;Accession: S77524  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1200 <KAN>  
A;Cross-references: UNIPROT:P73340; UNIPARC:UPI000003451; EMBL:D90905; GB:AB001339; NID  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C;Superfamily: chromosome segregation protein SMC1

Query Match 50.0%; Score 43; DB 2; Length 1200;  
Best Local Similarity 70.0%; Pred. No. 56;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 4 HYYLNGRTATR 13  
: | | | | |

Db 130 NYINGETAT 139

## RESULT 12

B90656  
IcmP-like protein [imported] - *Escherichia coli* (strain O157:H7, substrain RIMD 0509952)  
C;Species: *Escherichia coli*  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C;Accession: B90656  
R;Hayaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
gaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genom  
A;Reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: B90656  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1035 <HAY>  
A;Cross-references: UNIPROT:Q8X7W9; UNIPARC:UPI00001653A6; GB:BA000007; PIDN:BAB33641.1;  
A;Experimental source: strain O157:H7, substrain RIMD 0509952  
C;Genetics:  
A;Gene: ECs0218

Query Match 49.4%; Score 42.5; DB 2; Length 1035;  
Best Local Similarity 87.5%; Pred. No. 58;  
Matches 7; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2 WHHYLNG 9  
| | | | |  
| | | | |  
DB 329 WHHY-NG 335

## RESULT 13

C85507  
probable macrophage toxin [imported] - *Escherichia coli* (strain O157:H7, substrain EDL933;  
C;Species: *Escherichia coli*  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C;Accession: C85507  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: C85507  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1144 <STO>  
A;Cross-references: UNIPROT:Q8X7W9; UNIPARC:UPI000000D09DE; GB:AE005174; NID:gl2512953; P  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: Z0250

Query Match 49.4%; Score 42.5; DB 2; Length 1144;  
Best Local Similarity 87.5%; Pred. No. 64;  
Matches 7; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2 WHHYLNG 9  
| | | | |  
| | | | |  
DB 438 WHHY-NG 444

## RESULT 14

S36824  
beta-galactoside alpha-2,3-sialyltransferase (EC 2.4.99.4) - mouse  
C;Species: *Mus musculus* (house mouse)  
C;Date: 22-Jan-1994 #sequence\_revision 13-Mar-1997 #text\_change 05-Oct-2004  
C;Accession: S36824  
R;Lee, Y.C.; Kurosawa, N.; Hamamoto, T.; Nakaoka, T.; Tsuji, S.  
Eur. J. Biochem. 216, 377-385, 1993  
A;Title: Molecular cloning and expression of Gal-beta-1,3galNAc-alpha-2,3-sialyltransferase  
A;Reference number: S36824; MUID:93387288; PMID:8375377  
A;Accession: S36824  
A;Status: preliminary

```

A:Molecule type: mRNA
A:Residues: 1-337 <LEE>
A:Cross-references: UNIPROT:P54751; UNIPARC:UPI0000003F81; EMBL:X73523; NID:g402214; PID:
C:Superfamily: sialyltransferase
C:Keywords: glycosyltransferase

Query Match 48.8%; Score 42; DB 2; Length 337;
Best Local Similarity 54.5%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 WHYYLNGRTA 12
Db 294 WHYYWNPNSA 304

RESULT 15
154229
beta-galactoside alpha-2,3-sialyltransferase (EC 2.4.99.4) - human
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Oct-2004
C:Accession: I54229; A54898
R:Chang, M.L.; Eddy, R.L.; Shows, T.B.; Lau, J.T.
Glycobiology 5, 319-325, 1995
A:Title: Three genes that encode human beta-galactoside alpha 2,3-sialyltransferases. Sub
A:Reference number: I54229; MUID:95383839; PMID:7655169
A:Accession: I54229
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-340 <RES>
A:Cross-references: UNIPROT:Q11201; UNIPARC:UPI00000015E1; GB:L13972; NID:g410225; PIDN:
J. Kitagawa, H.; Paulson, J.C.
J. Biol. Chem. 269, 17872-17878, 1994
A:Title: Differential expression of five sialyltransferase genes in human tissues.
A:Reference number: A54898; MUID:94299495; PMID:8027041
A:Accession: A54898
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-11, 'V', 13-340 <KIT>
A:Cross-references: UNIPARC:UPI0000149058; GB:L29555; NID:g522196; PIDN:AAA36612.1; PID:
C:Genetics:
A:Gene: GDB:SIAT4A
A:Cross-references: GDB:384704
A:Map position: 3q21-3q28
C:Superfamily: sialyltransferase
C:Keywords: glycosyltransferase

Query Match 48.8%; Score 42; DB 2; Length 340;
Best Local Similarity 54.5%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 WHYYLNGRTA 12
Db 297 WHYYWNPNSA 307

RESULT 16
A45073
Gal beta 1,3GalNAc alpha 2,3-sialyltransferase - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 05-Oct-2004
C:Accession: A45073
R:Gillespie, W.; Kelm, S.; Paulson, J.C.
J. Biol. Chem. 267, 21004-21010, 1992
A:Title: Cloning and expression of the Gal beta 1, 3GalNAc alpha 2,3-sialyltransferase.
A:Reference number: A45073; MUID:93016016; PMID:1383214
A:Accession: A45073
A>Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-343 <GIL>
A:Cross-references: UNIPARC:UPI0000149053
A:Note: sequence extracted from NCBI backbone (NCBIN:116168, NCBI:P:116169)
C:Superfamily: sialyltransferase

```

GTP-binding protein ARD 1 - human  
N;Alternate names: ADP-ribosylation factor homolog  
C;Species: Homo sapiens (man)  
C;Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 02-Feb-2001  
C;Accession: A46054  
R;Mishima, K.; Tsuchiya, M.; Nightingale, M.S.; Moss, J.; Vaughan, M.  
J. Biol. Chem. 268, 8801-8807, 1993  
A;Title: ARD 1, a 64-kDa guanine nucleotide-binding protein with a carboxyl-terminal ADP  
A;Reference number: A46054; MUID: 93232038; PMID: 8473324  
A;Accession: A46054  
A;Status: preliminary  
A;Molecule type: nucleic acid  
A;Residues: 1-574 <MIS>  
A;Cross-references: UNIPARC:UPI000017C149  
A;Note: sequence inconsistent with the nucleotide translation  
A;Note: sequence extracted from NCBI backbone (NCBIN:129620, NCBIPI:129622)  
C;Genetics:  
A;Gene: GDB:ARD1; ARPD1  
A;Cross-references: GDB:139213; OMIM:601747  
A;Map position: 11q23.3-11q23.3  
C;Keywords: GTP binding; nucleotide binding; P-loop  
F:411-418/Region: nucleotide-binding motif A (P-loop)  
F:513-516/Region: GTP-binding NKXD motif

Query Match 48.8%; Score 42; DB 2; Length 574;  
Best Local Similarity 85.7%; Pred. No. 37;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLNL 8  
| | | | |  
DB 465 WKHYLYNL 471

RESULT 20  
S67483  
adenosinetriphosphatase 2 - malaria parasite (Plasmodium falciparum)  
N;Alternate names: ATPase 2  
C;Species: Plasmodium falciparum  
C;Date: 28-Oct-1996 #sequence\_revision 27-Feb-1997 #text\_change 09-Jul-2004  
C;Accession: S67483  
R;Trottein, F.; Cowman, A.F.  
Eur. J. Biochem. 227, 214-225, 1995  
A;Title: Molecular cloning and sequence of two novel P-type adenosinetriphosphatases frc  
A;Reference number: S67483; MUID: 95154293; PMID: 7851389  
A;Accession: S67483  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1553 <TRO>  
A;Cross-references: UNIPROT:Q9U421; UNIPARC:UPI000017B5ED; EMBL:U16955  
C;Genetics:  
A;Introns: 17/2

Query Match 48.8%; Score 42; DB 2; Length 1553;  
Best Local Similarity 77.8%; Pred. No. 1.1e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 HHYYLNGRT 11  
| | | | |  
DB 1345 HHYYFNIRT 1353

RESULT 21  
A61259  
glycoprotein S - porcine transmissible gastroenteritis virus (strain Miller) (fragments)  
N;Alternate names: E2 glycoprotein  
C;Species: porcine transmissible gastroenteritis virus  
C;Date: 12-May-1994 #sequence\_revision 27-Jun-1994 #text\_change 09-Jul-2004  
C;Accession: A61259; A33140; B33140  
R;Bae, I.; Jackwood, D.J.; Benfield, D.A.; Saif, L.J.; Wesley, R.D.; Hill, H.  
J. Clin. Microbiol. 29, 215-218, 1991  
A;Title: Differentiation of transmissible gastroenteritis virus from porcine respiratory  
he S glycoprotein gene.  
A;Reference number: A61259; MUID: 91131785; PMID: 1847152

A;Accession: A61259  
A;Status: not compared with conceptual translation  
A;Molecule type: genomic RNA  
A;Residues: 1-216 <BAE>  
A;Cross-references: UNIPROT:Q7LZU7; UNIPARC:UPI0000178664  
C;Superfamily: coronavirus E2 glycoprotein  
C;Keywords: glycoprotein

Query Match 47.7%; Score 41; DB 2; Length 216;  
Best Local Similarity 62.5%; Pred. No. 19;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLNG 9  
| | | | |  
DB 192 WGHFYNG 199

RESULT 22  
AB2349  
polysaccharide deacetylase [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp. PCC 7120  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 31-Dec-2004  
C;Accession: AB2349  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A;Reference number: AB1807; MUID: 21595285; PMID: 11759840  
A;Accession: AB2349  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-305 <KUR>  
A;Cross-references: UNIPROT:Q8YP55; UNIPARC:UPI000000CEACE; GB:BA0000019; PIDN:BAB76044.1;  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: all4345  
C;Superfamily: nodulation protein nodB

Query Match 47.7%; Score 41; DB 2; Length 305;  
Best Local Similarity 83.3%; Pred. No. 28;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHHY 6  
| | | | |  
DB 165 TWHHY 170

RESULT 23  
G82577  
phage-related integrase XF2288 [imported] - Xylella fastidiosa (strain 9a5c)  
C;Species: Xylella fastidiosa  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004  
C;Accession: G82577  
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc  
Nature 406, 151-157, 2000  
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A;Reference number: A82515; MUID: 20365717; PMID: 10910347  
A;Note: for a complete list of authors see reference number A59328 below  
A;Accession: G82577  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-353 <SIM>  
A;Cross-references: UNIPROT:Q9PB56; UNIPARC:UPI000000C29CA; GB:AE0004040; GB:AE0003849; NID:  
A;Experimental source: strain 9a5c  
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Brienes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.  
as-Neto, E.; Docena, C.; El-Dorri, H.; Paciniani, A.P.; Ferreira, A.J.S.  
submitted to Genbank, June 2000  
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigre  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.  
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;



F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A.; Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XP2288

Query Match 47.7%; Score 41; DB 2; Length 353;  
Best Local Similarity 60.0%; Pred. No. 32;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHYYLNGRT 11  
| : : : :  
Db 27 WESFYNGRT 36

RESULT 24  
T41390  
zinc finger protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 05-Oct-2004  
C:Accession: T41390  
R;Murphy, L.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, May 1998  
A:Reference number: Z21990  
A:Accession: T41390  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-463 <MUR>  
A:Cross-references: UNIPROT:O59811; UNIPARC:UPI000006C562; EMBL:AL023592; PIDN:CAA19119.  
A:Experimental source: strain 972h-; cosmid c550  
C:Genetics:  
A:Gene: SPDB:SPCC550.15C  
A:Map position: 3  
C:Superfamily: zinc finger protein

Query Match 47.7%; Score 41; DB 2; Length 463;  
Best Local Similarity 50.0%; Pred. No. 43;  
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 WHYYLNGRTAT 13  
| : : : :  
Db 28 WHYYLKRKVAS 39

RESULT 25  
E81182  
hypothetical protein NMB0570 [imported] - Neisseria meningitidis (strain MC58 serogroup  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C:Accession: E81182  
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755; PMID:10710307  
A:Accession: E81182  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-473 <TET>  
A:Cross-references: UNIPROT:Q9K0M2; UNIPARC:UPI00000C44F5; GB:AE002413; GB:AE002098; NID  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB0570

Query Match 47.7%; Score 41; DB 2; Length 473;  
Best Local Similarity 83.3%; Pred. No. 44;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHHY 6  
| : : : :  
Db 51 TWHHY 56

RESULT 26  
B81919  
probable membrane protein NMA0753 [imported] - Neisseria meningitidis (strain Z2491 sero  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 09-Jul-2004  
C:Accession: B81919  
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell  
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: A81775; MUID:20222556; PMID:10761919  
A:Accession: B81919  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-473 <PAR>  
A:Cross-references: UNIPROT:Q9JVP7; UNIPARC:UPI00000C4A32; GB:AL162754; GB:AL157959; NID:  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA0753

Query Match 47.7%; Score 41; DB 2; Length 473;  
Best Local Similarity 83.3%; Pred. No. 44;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWHHY 6  
| : : : :  
Db 51 TWHHY 56

RESULT 27  
S24284  
E2 glycoprotein precursor - porcine respiratory virus (strain 86/137004)  
N:Alternate names: peplomer glycoprotein; spike glycoprotein  
C:Species: porcine respiratory virus  
A:Variety: strain 86/137004  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004  
C:Accession: S24284; S21871  
R;Britton, P.; Mawditt, K.L.; Page, K.W.  
Virus Res. 21, 181-198, 1991  
A:Title: The cloning and sequencing of the virion protein genes from a British isolate of  
A:Reference number: S24279; MUID:92116634; PMID:1662846  
A:Accession: S24284  
A:Molecule type: genomic RNA  
A:Residues: 1-1225 <BRI>  
A:Cross-references: UNIPROT:P27655; UNIPARC:UPI0000138673; EMBL:X60089; NID:G58983; PIDN:  
A:Experimental source: strain 86/137004  
C:Superfamily: coronavirus E2 glycoprotein  
C:Keywords: Glycoprotein; transmembrane protein  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-1225/Product: E2 glycoprotein #status predicted <EG>  
F:1167-1187/Domain: transmembrane #status predicted <TM>  
F:26,61,110,121,138,151,181,225,256,292,308,330,338,370,480,501,556,595,610,616,697,850,9  
Query Match 47.7%; Score 41; DB 1; Length 1225;  
Best Local Similarity 62.5%; Pred. No. 1.2e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLNG 9  
| : : : :  
Db 203 WGHFYNG 210

RESULT 28  
A36607  
E2 glycoprotein - porcine respiratory virus (strain RM4)  
C:Species: porcine respiratory virus  
C:Date: 30-Aug-1991 #sequence\_revision 30-Aug-1991 #text\_change 09-Jul-2004  
C:Accession: A36607

R;Rasschaert, D.; Duarte, M.; Laude, H.  
J. Gen. Virol. 71, 2599-2607, 1990  
A;Title: Porcine respiratory coronavirus differs from transmissible gastroenteritis virus  
A;Reference number: A36607; MUID:91073120; PMID:2174956  
A;Accession: A36607  
A;Status: preliminary  
A;Molecule type: genomic RNA  
A;Residues: 1-1225 <RAS>  
A;Cross-references: UNIPROT:P24413; UNIPARC:UPI0000138674; GB:224675; NID:g395057; PIDN:  
C;Superfamily: coronavirus E2 glycoprotein  
C;Keywords: glycoprotein; transmembrane protein

Query Match 47.7%; Score 41; DB 2; Length 1225;  
Best Local Similarity 62.5%; Pred. No. 1.2e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLNG 9  
| | : | : |  
Db 203 WGHFYNG 210

RESULT 29  
VGTHE3  
E2 glycoprotein precursor - porcine transmissible gastroenteritis virus (strain Purdue)  
N;Alternate names: peplomer protein; spike glycoprotein  
C;Species: porcine transmissible gastroenteritis virus  
C;Date: 30-Jun-1990 #sequence\_revision 30-Jun-1990 #text\_change 12-Apr-1996  
C;Accession: JS0336  
R;Jacobs, L.; de Groot, R.; van der Zeijst, B.A.M.; Horzinek, M.C.; Spaan, W.  
Virus Res. 8, 363-371, 1987  
A;Title: The nucleotide sequence of the peplomer gene of porcine transmissible gastroent  
s (FIPV).  
A;Reference number: JS0336; MUID:88129049; PMID:2829461  
A;Accession: JS0336  
A;Molecule type: mRNA  
A;Residues: 1-1447 <JAC>  
A;Cross-references: UNIPARC:UPI0000174A62  
C;Superfamily: coronavirus E2 glycoprotein  
C;Keywords: glycoprotein; transmembrane protein  
F;1-16/Domain: signal sequence #status predicted <SIG>  
F;17-1447/Product: E2 glycoprotein #status predicted <MAT>  
F;1387-1431/Domain: transmembrane #status predicted <TMN>  
F;26,42,71,94,243,250,285,334,345,362,403,447,514,530,552,592,702,723,778,817,832,838,91

Query Match 47.7%; Score 41; DB 1; Length 1447;  
Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLNG 9  
| | : | : |  
Db 425 WGHFYNG 432

RESULT 30  
VGTHE2  
E2 glycoprotein precursor - porcine transmissible gastroenteritis virus (strain Purdue-1  
N;Alternate names: spike glycoprotein  
C;Species: porcine transmissible gastroenteritis virus  
A;Variety: strain Purdue-115  
C;Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 09-Jul-2004  
C;Accession: A27106; S01738  
R;Rasschaert, D.; Laude, H.  
J. Gen. Virol. 68, 1883-1890, 1987  
A;Title: The predicted primary structure of the peplomer protein E2 of the porcine coron  
A;Reference number: A27106; MUID:87253116; PMID:3037011  
A;Accession: A27106  
A;Molecule type: genomic RNA  
A;Residues: 1-1447 <RAS>  
A;Cross-references: UNIPROT:P07946; UNIPARC:UPI0000138672; GB:X05695; GB:D00118; NID:g59  
R;Rasschaert, D.; Geifli, J.; Laude, H.  
Biochimie 69, 591-600, 1987  
A;Title: Enteric coronavirus TGEV: partial sequence of the genomic RNA, its organization

A;Reference number: S01738; MUID:88078100; PMID:2825819  
A;Accession: S01738  
A;Molecule type: genomic RNA  
A;Residues: 1434-1447 <RAW>  
A;Cross-references: UNIPARC:UPI0000174A6D; EMBL:X06371  
A;Experimental source: strain Purdue-115  
C;Superfamily: coronavirus E2 glycoprotein  
C;Keywords: glycoprotein; transmembrane protein  
F;1-16/Domain: signal sequence #status predicted <SIG>  
F;17-1447/Product: E2 glycoprotein #status predicted <MAT>  
F;1387-1431/Domain: transmembrane #status predicted <TMN>  
F;26,42,71,94,243,250,285,334,345,362,403,447,514,530,552,592,702,723,778,817,832,838,91

Query Match 47.7%; Score 41; DB 1; Length 1447;  
Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLNG 9  
| | : | : |  
Db 425 WGHFYNG 432

RESULT 31  
A43573  
E2 glycoprotein precursor - porcine transmissible gastroenteritis virus (strain Miller)  
N;Alternate names: peplomer glycoprotein; spike glycoprotein  
C;Species: porcine transmissible gastroenteritis virus  
C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004  
C;Accession: A43573  
R;Wesley, R.D.  
Adv. Exp. Med. Biol. 276, 301-306, 1990  
A;Title: Nucleotide sequence of the E2-peplomer protein gene and partial nucleotide sequ  
A;Reference number: A43573; MUID:91353366; PMID:1966416  
A;Accession: A43573  
A;Molecule type: genomic RNA  
A;Residues: 1-1449 <WES>  
A;Cross-references: UNIPROT:P33470; UNIPARC:UPI0000174A6F; GB:S51223; NID:g234109; PIDN:  
C;Note: the authors translated the codon GAA for residue 388 as Cys  
C;Superfamily: coronavirus E2 glycoprotein  
C;Keywords: glycoprotein; transmembrane protein  
F;1-16/Domain: signal sequence #status predicted <SIG>  
F;17-1449/Product: E2 glycoprotein #status predicted <B2G>  
F;1027-1043/Region: hydrophobic  
F;1391-1411/Domain: transmembrane #status predicted <TMN>  
F;26,42,71,94,243,250,285,334,345,362,375,405,449,516,532,554,594,704,725,780,819,834,84  
d

Query Match 47.7%; Score 41; DB 1; Length 1449;  
Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLNG 9  
| | : | : |  
Db 427 WGHFYNG 434

RESULT 32  
VGTHE5  
E2 glycoprotein precursor - porcine transmissible gastroenteritis virus (strain FS772/70)  
N;Alternate names: peplomer glycoprotein; spike glycoprotein  
C;Species: porcine transmissible gastroenteritis virus  
C;Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 09-Jul-2004  
C;Accession: B43489; S11728  
R;Britton, P.; Page, K.W.  
Virus Res. 18, 71-80, 1990  
A;Title: Sequence of the S gene from a virulent British field isolate of transmissible g  
A;Reference number: A43489; MUID:91188698; PMID:1964522  
A;Accession: B43489  
A;Molecule type: mRNA  
A;Residues: 1-1449 <BRI>  
A;Cross-references: UNIPROT:P18450; UNIPARC:UPI000013866F; GB:X53128; NID:g61377; PIDN:C  
C;Superfamily: coronavirus E2 glycoprotein  
C;Keywords: glycoprotein; transmembrane protein

F;1-16/Domain: signal sequence #status predicted <SIG>  
F;17-1449/Product: E2 glycoprotein #status predicted <E2G>  
F;1027-1043/Region: hydrophobic  
F;1395-1411/Domain: transmembrane #status predicted <TMN>  
F;26,42,71,94,243,250,285,334,345,362,375,405,449,516,532,554,594,704,725,780,819,834,84  
d

Query Match 47.7%; Score 41; DB 1; Length 1449;  
Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLNG 9  
| | | | |  
Db 427 WGHFYNG 434

RESULT 33  
E2 glycoprotein precursor - porcine transmissible gastroenteritis virus  
N;Alternate names: envelope protein; spike protein  
C;Species: porcine transmissible gastroenteritis virus  
C;Date: 23-Nov-1994 #sequence\_revision 19-Apr-1996 #text\_change 09-Jul-2004  
C;Accession: S65851; S47423  
R;Chen, C.M.; Cavanagh, D.; Britton, P.  
Virus Res. 38, 83-89, 1995  
A;Title: Cloning and sequencing of a 8.4-kb region from the 3'-end of a Taiwanese viru  
A;Reference number: S65850; MUID:96060227; PMID:8546012  
A;Accession: S65851  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: genomic RNA  
F;1-1449 <CH2>  
A;Residues: 1-1449  
A;Cross-references: UNIPROT:Q88510; UNIPARC:UPI00000F855B; EMBL:Z35758; NID:G529246; PID  
A;Experimental source: Taiwanese field isolate  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994  
C;Genetics:  
A;Gene: S  
C;Superfamily: coronavirus E2 glycoprotein  
C;Keywords: glycoprotein; transmembrane protein  
F;1-16/Domain: signal sequence #status predicted <SIG>  
F;17-1449/Product: E2 glycoprotein #status predicted <E2G>

Query Match 47.7%; Score 41; DB 2; Length 1449;  
Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLNG 9  
| | | | |  
Db 427 WGHFYNG 434

RESULT 34  
E2 glycoprotein precursor - canine coronavirus (strain Inseavc-1)  
N;Alternate names: spike glycoprotein  
C;Species: canine coronavirus  
C;Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 09-Jul-2004  
C;Accession: JQ1719  
R;Horsburgh, B.C.; Brierley, I.; Brown, T.D.K.  
J. Gen. Virol. 73, 2849-2862, 1992  
A;Title: Analysis of a 9.6 kb sequence from the 3' end of canine coronavirus genomic RNA  
A;Reference number: PQ0481; MUID:93057357; PMID:1431811  
A;Accession: JQ1719  
A;Molecule type: genomic RNA  
A;Residues: 1-1451 <HOR>  
A;Cross-references: UNIPROT:P36300; UNIPARC:UPI0000138669; DDBJ:D13096; NID:G406193; PID  
C;Genetics:  
A;Gene: S  
C;Superfamily: coronavirus E2 glycoprotein  
C;Keywords: glycoprotein; transmembrane protein  
F;1-18/Domain: signal sequence #status predicted <SIG>  
F;19-1451/Product: spike glycoprotein #status predicted <NAT>  
F;1394-1412/Domain: transmembrane #status predicted <TMN>  
F;28,66,94,142,175,209,235,242,289,338,349,366,379,409,453,520,536,557,707,728,783,821,8

d

Query Match 47.7%; Score 41; DB 1; Length 1451;  
Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLNG 9  
| | | | |  
Db 431 WGHFYNG 438

RESULT 35  
VG1H79  
E2 glycoprotein precursor - feline infectious peritonitis virus (strain 79-1146)  
N;Alternate names: peplomer glycoprotein; spike glycoprotein  
C;Species: feline infectious peritonitis virus  
C;Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 09-Jul-2004  
C;Accession: A27171  
R;De Groot, R.J.; Maduro, J.; Lenstra, J.A.; Horzinek, M.C.; Van Der Zeijst, B.A.M.; Spa  
J. Gen. Virol. 68, 2639-2646, 1987  
A;Title: cDNA cloning and sequence analysis of the gene encoding the peplomer protein of  
A;Reference number: A27171; MUID:88034948; PMID:3312491  
A;Accession: A27171  
A;Molecule type: genomic RNA  
A;Residues: 1-1452 <DEG>  
A;Cross-references: UNIPROT:P10033; UNIPARC:UPI000005FIAB; GB:X06170; GB:D00150; NID:G58;  
C;Superfamily: coronavirus E2 glycoprotein  
C;Keywords: glycoprotein; transmembrane protein  
F;1-19/Domain: signal sequence #status predicted <SIG>  
F;1-19/Domain: transmembrane #status predicted <TM1>  
F;20-1452/Product: spike glycoprotein #status predicted <SPG>  
F;1394-1414/Domain: transmembrane #status predicted <TM2>  
F;29,95,174,208,234,241,288,337,348,365,408,452,483,519,535,557,565,707,728,783,822,837,8  
ted

Query Match 47.7%; Score 41; DB 1; Length 1452;  
Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLNG 9  
| | | | |  
Db 430 WGHFYNG 437

RESULT 36  
S41453  
spike protein - canine coronavirus  
C;Species: canine coronavirus  
C;Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 09-Jul-2004  
C;Accession: S41453  
R;Wesseling, J.G.; Vennema, H.; Godeke, G.J.; Spaan, W.J.M.; Horzinek, M.C.; Rottier, P.  
submitted to the EMBL Data Library, December 1993  
A;Description: Nucleotide sequence and expression of the spike (S) gene of canine corona  
A;Reference number: S41453  
A;Accession: S41453  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-1453 <WES>  
A;Cross-references: UNIPROT:Q65984; UNIPARC:UPI00000F580F; EMBL:X77047; NID:G452379; PID  
C;Superfamily: coronavirus E2 glycoprotein

Query Match 47.7%; Score 41; DB 2; Length 1453;  
Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WHYYLNG 9  
| | | | |  
Db 431 WGHFYNG 438

RESULT 37  
AG2106  
hypothetical protein alr2406 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C;Accession: AG2106  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Tanaka, K.; Shimizu, T.; Shiraiwa, Y.; Kusuda, M.; Tabata, S.; Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
A;Reference number: AB1807; PMID:21595285; PMID:11759840  
A;Accession: AG2106  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-250 <KUR>  
A;Cross-references: UNIPROT:Q44510; UNIPARC:UP10000013BBED; GB:BA000019; PIDN:BAE74105.1;  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: alr2406

```

Query Match      47.1%; Score 40.5; DB 2; Length 250;
Best Local Similarity 57.1%; Pred. No. 27;
Matches 8; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Qy      1 TWH-HYYLNGRTAT 13
      ||| |||| : |
      ||| |||| : |
Db      104 TWHAHYYLQRLFT 117

RESULT 38
B88088
protein B0454.2 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C;Accession: B88088
R;anonymous, The C. elegans Sequencing Consortium.

```

A:Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biological Science 282, 2012-2016, 1998  
 A:Reference number: A75000; PMID:99069613; PMID:9851916  
 A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.elegans/ Science 282, 2012-2016, 1998  
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 283, 2103, 1999  
 A:Accession: B88088  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-324 <STO>  
 A:Cross-references: UNIPROT:O17172; UNIPARC:UPI00000797C5; GB:chr\_II; PIDN:AB70940.1; PIR:AB70940.1  
 C:Genetics:  
 A:Gene: B0454.2  
 A:Map position: 2

Query Match	47.1%	Score 40.5;	DB 2;	Length 324;
Best Local Similarity	50.0%;	Pred. No. 36;		
Matches	7;	Conservative	2;	Mismatches
			2;	Indels
			3;	Gaps

Qy 2 WH-- --HYLNGRTA 12  
Db 11 WHLIYHYISGTIA 24

RESULT 39  
D81173  
hypothetical protein NMB0649 [imported] - Neisseria meningitidis (strain MC58 serogroup  
C;Species: Neisseria meningitidis  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C;Accession: D81173  
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
xi, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A;Reference number: AB1000; MUID:20175755; PMID:10710307  
A;Accession: D81173  
A;Status: preliminary  
A;Molecule type: DNA

A;Residues: 1-74 <TET>  
A;Cross-references: UNIPROT:Q9K0F7; UNIPARC:UPI000004460F; GB:AE002420; GB:AE002098; NID  
A;Experimental source: serogroup B, strain MC58  
C;Genetics:  
A;Gene: NMB0649

Query Match	46.5%	Score 40;	DB 2;	Length 74;
Best Local Similarity	60.0%;	Pred. No. 8.9;		
Matches	6;	Conservative	1;	Mismatches 3; Indels 0; Gaps 0;
Qy	1	TWHHYLYNGR	10	
Db	41	TWHHODTGR	50	

## RESULT 40

C81931  
hypotheical protein NMA0855 [imported] - Neisseria meningitidis (strain Z2491 serogroup  
C/Species: Neisseria meningitidis  
C/Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 09-Jul-2004  
C/Accession: C81931  
F:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
; Holroyd, S.; Jagals, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A/Reference number: A81775; MUID:20222556; PMID:10761919  
A/Accession: C81931  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-81 <PAR>  
A/Cross-references: UNIPROT:Q9JVF9; UNIPARC:UPI00000C4A77; GB:AL157959; NID:  
A/Experimental source: serogroup A, strain Z2491  
C/Genetics:  
A/Gene: NMA0855

Query Match 46.5%; Score 40; DB 2; Length 81;  
Best Local Similarity 60.0%; Pred. No. 9.8;  
Matches 6; Conservative 1; Mismatches 3; Indels

Qy 1 TWHHYLNGR 10  
|||: ||  
Db 48 TWHHHQDTGR 57

## RESULT 41

T21783  
hypothetical protein F35E2.3 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T21783  
R:Lennard, N.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19471  
A:Accession: T21783  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-227 <WIL>  
A/Cross-references: UNIPROT:O62229; UNIPARC:UPI0000077B3E; EMBL:Z81528; PDB:1D1D  
A:Experimental source: clone F35E2

A;Map position: 1  
A: Introns: 42/1: 102/3: 175/1

Query Match	46.5%	Score 40;	DB 2;	Length 227;
Best Local Similarity	54.5%;	Pred. No. 30;		
Matches 6;	Conservative	3;	Mismatches	2;
			Indels	0;
			Gaps	0

QY 1 TWHYYLNGRT 11  
|:|::||  
Db 207 TYSHYYVDSRT 217

RESULT 42  
AD1209  
N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) lmo1076 - Listeria monocytogenes (strain  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A:Title: Comparative Genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AD1209  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-572 <GLA>  
A:Cross-references: UNIPROT:Q8Y842; UNIPARC:UPI0000054CB2; GB:NC\_003210; PIDN:CAC99154.1  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo1076  
C:Keywords: hydrolase

Query Match 46.5%; Score 40; DB 2; Length 572;  
Best Local Similarity 45.5%; Pred. No. 80;  
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TWHHYYLNGRT 11  
||: : |||:  
Db 472 TWYQFQVNGKT 482

RESULT 43  
T08984  
auxin response factor 7 homolog F6G3.110 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 22-Oct-1999  
C:Accession: T08984  
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X  
submitted to the Protein Sequence Database, May 1999  
A:Reference number: Z16520  
A:Accession: T08984  
A:Molecule type: DNA  
A:Residues: 1-653 <BEV>  
A:Cross-references: UNIPARC:UPI0000A0314; EMBL:AL078464; GSPDB:GN00062; ATSP:F6G3.110  
A:Experimental source: cultivar Columbia; BAC clone F6G3  
C:Genetics:  
A:Gene: ATSP:F6G3.110  
A:Map position: 4  
A:Introns: 362/3

Query Match 46.5%; Score 40; DB 2; Length 653;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HHYYLN 8  
|||||  
Db 464 HHYYLN 469

RESULT 44  
T37276  
protein PEX6 homolog - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T37276; T03904; T37307  
R:Kamiryo, T.; Bun-ya, M.  
submitted to the EMBL Data Library, February 1998  
A:Description: C.elegans cDNA coding PEX6 homologue.  
A:Reference number: Z21663  
A:Accession: T37276  
A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA  
A:Residues: 1-720 <KNA>  
A:Cross-references: UNIPROT:O16270; UNIPARC:UPI000007F9CB; EMBL:AB010968; PIDN:BAA33544.1  
R:Du, Z.; Le, T.T.; Holmes, A.  
submitted to the EMBL Data Library, July 1997  
A:Description: The sequence of C. elegans cosmid F39G3.  
A:Reference number: Z15131  
A:Accession: T03904  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-720 <DUZ>  
A:Cross-references: UNIPARC:UPI000007F9CB; EMBL:AF016424; NID:G2291203; PIDN:AAB65332.1.  
R:Kamiryo, T.; Bun-ya, M.; Chenea, S.  
submitted to the EMBL Data Library, March 1998  
A:Description: Genomic DNA sequence of PEX6 homolog of C. elegans.  
A:Reference number: Z21673  
A:Accession: T37307  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-720 <KA2>  
A:Cross-references: UNIPARC:UPI000007F9CB; EMBL:AB012224; PIDN:BAA76440.1  
C:Genetics:  
A:Map position: V  
A:Introns: 47/2; 122/3; 170/2; 338/1; 458/3; 659/3  
A:Note: F39G3.7

Query Match 46.5%; Score 40; DB 2; Length 720;  
Best Local Similarity 58.3%; Pred. No. 1e+02;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TWHHYYLNGRTA 12  
||| ||||: |  
Db 393 TWQYYLNEKLA 404

RESULT 45  
E97177  
uncharacterized conserved membrane protein, affecting LPS biosynthesis CAC2251 [imported]  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 05-Oct-2004  
C:Accession: E97177  
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: E97177  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-723 <KUR>  
A:Cross-references: UNIPROT:Q97GW4; UNIPARC:UPI00000CA479; GB:AE001437; PIDN:AAK80208.1;  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC2251  
C:Superfamily: uncharacterized conserved protein

Query Match 46.5%; Score 40; DB 2; Length 723;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TWHHYYL 7  
||| |||  
Db 416 TWHPYYL 422

RESULT 46  
F88448  
protein C45G9.10 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: F88448  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998

A>Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biological  
 A/Reference number: A75000; MUID:99069613; PMID:9851916  
 A/Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
 A/Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
 A/Accession: F88448  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-971 <STO>  
 A/Cross-references: UNIPROT:Q09281; UNIPARC:UPI000013BF17; GB:chr\_III; PIDN:AAA62553.1;  
 C/Genetics:  
 A/Gene: C45G9.10  
 A/Map position: 3

Query Match 46.5%; Score 40; DB 2; Length 971;  
 Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 HHYYLNGRT 11  
 |||||  
 Db 696 HKYYINGET 704

RESULT 47  
 S46837  
 hypothetical protein YHL023c - yeast (*Saccharomyces cerevisiae*)  
 C/Species: *Saccharomyces cerevisiae*  
 C/Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 09-Jul-2004  
 C/Accession: S46837  
 R/Favell, T.  
 submitted to the EMBL Data Library, June 1994  
 A/Description: The sequence of *S. cerevisiae* cosmid 9433.  
 A/Reference number: S46796  
 A/Accession: S46837  
 A/Molecule type: DNA  
 A/Residues: 1-1146 <FAV>  
 A/Cross-references: UNIPROT:P38742; UNIPARC:UPI000013B1D0; EMBL:U11582; NID:g2289793; PI  
 C/Genetics:  
 A/Gene: MIPS:YHL023c  
 A/Cross-references: SGD:S0001015  
 A/Map position: 8L

Query Match 46.5%; Score 40; DB 2; Length 1146;  
 Best Local Similarity 46.2%; Pred. No. 1.7e+02;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 WHYYLNGTATR 14  
 :|||:|  
 Db 343 YHHYHKNATSOR 355

RESULT 48  
 H82802  
 fibrial assembly protein XF0478 [imported] - *Xylella fastidiosa* (strain 9a5c)  
 C/Species: *Xylella fastidiosa*  
 C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004  
 C/Accession: H82802  
 R/anonymos, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
 Nature 406, 151-157, 2000  
 A/Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
 A/Reference number: A82515; MUID:20365717; PMID:10910347  
 A/Note: for a complete list of authors see reference number A59328 below  
 A/Accession: H82802  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1472 <SIM>  
 A/Cross-references: UNIPROT:Q9PG24; UNIPARC:UPI00000C2413; GB:AE003897; GB:AE003849; NII  
 A/Experimental source: strain 9a5c  
 R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.  
 A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A/Reference number: A59328  
 A/Contents: annotation  
 C/Genetics:  
 A/Gene: XF0478

Query Match 46.5%; Score 40; DB 2; Length 1472;  
 Best Local Similarity 62.5%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 HHYYLNGR 10  
 |||||  
 Db 1032 HHYYVDGK 1039

RESULT 49  
 E83765  
 hypothetical protein BH0925 [imported] - *Bacillus halodurans* (strain C-125)  
 C/Species: *Bacillus halodurans*  
 C/Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004  
 C/Accession: E83765  
 R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiran  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A/Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
 A/Reference number: A83650; MUID:20512582; PMID:11058132  
 A/Accession: E83765  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-108 <STO>  
 A/Cross-references: UNIPROT:Q9KEC8; UNIPARC:UPI00000C3A2F; GB:AP001510; GB:BA000004; NID:  
 A/Experimental source: strain C-125  
 C/Genetics:  
 A/Gene: BH0925  
 C/Superfamily: *Bacillus subtilis* hypothetical protein yfhh

Query Match 45.3%; Score 39; DB 2; Length 108;  
 Best Local Similarity 66.7%; Pred. No. 20;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTA 12  
 :|||  
 Db 71 TFHISYLNGRFA 82

RESULT 50  
 KLCHI  
 calcium-binding protein, vitamin D-dependent - chicken  
 N/Alternate names: avian type CaBP; calbindin; large CaBP  
 C/Species: *Gallus gallus* (chicken)  
 C/Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 09-Jul-2004  
 C/Accession: A40926; A93605; A94129; A29957; A94166; A03070; A28026; JT0298  
 R/Minghetti, P.P.; Cancala, L.; Fujisawa, Y.; Theofan, G.; Norman, A.W.  
 Mol. Endocrinol. 2, 355-367, 1988  
 A/Title: Molecular structure of the chicken vitamin D-induced calbindin-D-28K gene reveal  
 A/Reference number: A40926; MUID:88246474; PMID:2967915  
 A/Accession: A40926  
 A/Molecule type: DNA  
 A/Residues: 1-262 <MIN>  
 A/Cross-references: UNIPROT:P04354; UNIPARC:UPI00001712BA; GB:M31139  
 R/Wilson, P.W.; Harding, M.; Lawson, D.E.M.  
 Nucleic Acids Res. 13, 8867-8881, 1985  
 A/Title: Putative amino acid sequence of chick calcium-binding protein deduced from a con  
 A/Reference number: A93605; MUID:86093684; PMID:3841205  
 A/Accession: A93605  
 A/Molecule type: mRNA  
 A/Residues: 1-262 <WTL>  
 A/Cross-references: UNIPARC:UPI00001712BA; GB:X03343; NID:g63170; PIDN:CAA27049.1; PID:g  
 A/Note: the authors translated the codon ACU for residue 70 as Tyr and GAU for residues 1

R;Hunziker, W.  
 Proc. Natl. Acad. Sci. U.S.A. 83, 7578-7582, 1986  
 A;Title: The 28-kDa vitamin D-dependent calcium-binding protein has a six-domain structure  
 A;Reference number: A94129; MUID:87016992; PMID:3463988  
 A;Accession: A94129  
 A:Molecule type: mRNA  
 A:Residues: 1-262 <HUN>  
 A:Cross-references: UNIPARC:UPI00001712BA; GB:MI4230; NID:G211429; PIDN:AAA48659.1; PID:  
 R;Wilson, P.W.; Rogers, J.; Harding, M.; Pohl, V.; Pattyn, G.; Lawson, D.E.M.  
 J. Mol. Biol. 200, 615-625, 1988  
 A;Title: Structure of chick chromosomal genes for calbindin and calretinin.  
 A;Reference number: A29957; MUID:88316929; PMID:3411606  
 A;Accession: A29957  
 A:Molecule type: DNA  
 A:Residues: 1-262 <WI2>  
 A:Cross-references: UNIPARC:UPI00001712BA; EMBL:X06629; NID:G63133; PIDN:CAA29843.1; PID:  
 R;Fullmer, C.S.; Wasserman, R.H.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 4772-4776, 1987  
 A;Title: Chicken intestinal 28-kilodalton calbindin-D: complete amino acid sequence and  
 A;Reference number: A94166; MUID:87260872; PMID:3474624  
 A;Accession: A94166  
 A:Molecule type: protein  
 A:Residues: 2-262 <FUL>  
 A:Cross-references: UNIPARC:UPI0000126D72  
 C;Comment: This protein is involved in vitamin D-stimulated calcium absorption from the  
 D; without it, synthesis quickly stops.  
 C;Comment: Four calcium ions can be bound per molecule.  
 C;Comment: The six domains of this protein are thought to derive from duplication of an  
 -86 and 231-260, respectively) no longer bind calcium.  
 C;Genetics:  
 A;Introns: 28/1; 53/3; 78/3; 106/3; 125/3; 151/3; 170/2; 183/3; 201/3; 225/3  
 C;Superfamily: calretinin; calmodulin repeat homology  
 C;Keywords: blocked amino end; calcium binding; duplication; EF hand; vitamin D  
 F;2-262/Product: calcium-binding protein, vitamin D-dependent #status experimental <MAT>  
 F;12-44/Domain: calmodulin repeat homology <EF1>  
 F;54-86/Domain: calmodulin repeat homology <EF2>  
 F;99-131/Domain: calmodulin repeat homology <EF3>  
 F;143-175/Domain: calmodulin repeat homology <EF4>  
 F;187-219/Domain: calmodulin repeat homology <EF5>  
 F;2/Modified site: blocked amino end (Thr) (in mature form) #status experimental  
 F;25-27, 29, 31, 36/Binding site: calcium (Asp, Asn, Tyr, Glu) #status predicted  
 F;112,114,116,118,123/Binding site: calcium (Asp, Asp, Ser, Phe, Glu) #status predicted  
 F;156,158,160,162,167/Binding site: calcium (Asp, Asn, Lys, Glu) #status predicted  
 F;200,202,204,206,211/Binding site: calcium (Asp, Asp, Asn, Tyr, Glu) #status predicted

Query Match 45.3%; Score 39; DB 1; Length 262;  
 Best Local Similarity 40.0%; Pred.No. 51;  
 Matches 6; Conservative 3; Mismatches 0; Indels 6; Gaps 1;

QY 2 WHRY-----YLNGR 10  
 |||||  
 |||||  
 Db 21 WHHYDSGNGYMDGK 35

Search completed: June 5, 2006, 12:53:54  
 Job time : 25.5342 secs



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Biocollaboration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:32:17 ; Search time 122.74' Seconds

(without alignments)  
105.510 Million cell updates/sec

Title: US-10-645-659A-10

Perfect score: 86

Sequence: 1 TWHYYLNGRTATR 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : Uniprot 7.2.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	86	100.0	543	1	HPSE_HUMAN
2	79	91.9	523	1	HPSE_CHICK
3	78	90.7	536	1	HPSE_RAT
4	77	89.5	535	1	HPSE_MOUSE
5	77	89.5	574	2	Q333X6_SPALAX
6	77	89.5	574	2	Q333X7_GRODE
7	77	89.5	574	2	Q333X8_GRODE
8	77	89.5	574	2	Q333X9_GRODE
9	74	86.0	545	1	HPSE_BOVIN
10	66	76.7	255	2	Q4TGC8_TETNG
11	66	76.7	533	2	Q4SYR6_TETNG
12	55	64.0	535	2	Q8T108_BOMMO
13	52	60.5	532	2	Q9PTB2_ICTFU
14	52	60.5	597	2	Q4TB80_TETNG
15	51	59.3	1179	2	O59462_PYROCO
16	50	58.1	124	2	Q4WRV8_ASPTU
17	49	57.0	52	2	Q2KXK8_BORAV
18	47	54.7	138	2	Q7P1C9_CHRVO
19	47	54.7	215	2	O5C4P3_SCHJA
20	47	54.7	382	2	Q7VJ45_HELHP
21	47	54.7	383	2	Q41029_GIBZE
22	47	54.7	1159	1	KCMAL1_CANFA
23	46	53.5	64	2	Q8BSE9_MOUSE
24	46	53.5	186	2	Q418T2_GIBZE
25	46	53.5	190	2	Q862K9_PODAN
26	46	53.5	208	2	Q2UL82_ASPOR
27	46	53.5	213	2	Q7S4E2_NEUCR
28	46	53.5	213	2	Q5B387_EMENI
29	46	53.5	279	2	Q4WXL6_ASPTU
30	46	53.5	301	2	Q5D288_VIBF1
31	46	53.5	349	2	Q8BSA0_MOUSE

32	46	53.5	349	2	Q5ZJJ0_CHICK	Q5ZJJ0 gallus gall
33	46	53.5	349	2	Q70D58_CHICK	Q70D58 gallus gall
34	46	53.5	350	1	SIA4B_HUMAN	Q16842 h cmp-n-ace
35	46	53.5	350	1	SIA4B_PANTR	Q6K558 p cmp-n-ace
36	46	53.5	350	1	SIA4B_RAT	Q11205 r cmp-n-ace
37	46	53.5	350	2	Q6H8M9_BOVIN	Q6H8M9 bos taurus
38	46	53.5	350	2	Q8BPL0_MOUSE	Q8BPL0 m o day neo
39	46	53.5	350	2	Q91WH6_MOUSE	Q91WH6 mus musculus
40	46	53.5	351	2	Q6NU41_XENLA	Q6NU41 xenopus lae
41	46	53.5	892	2	Q4AXU3_9BURK	Q4AXU3 polaromonas
42	45	52.3	190	2	O80438_ARATH	O80438 arabidopsis
43	45	52.3	448	2	Q2URZ0_ASPOR	Q2URZ0 aspergillus
44	45	52.3	558	2	Q333X5_SPALAX	Q333X5 spalax juda
45	45	52.3	612	2	Q4PHG9_TETNG	Q4PHG9 tetraodon n
46	44.5	51.7	3050	2	Q6PY34_9POTY	Q6PY34 hordeum mos
47	44	51.2	161	2	Q7V6T4_PROMM	Q7V6T4 prochloroco
48	44	51.2	234	2	Q4SB93_TETNG	Q4SB93 tetraodon n
49	44	51.2	236	2	Q404U7_9ROB	Q404U7 jannaschia
50	44	51.2	260	2	Q5W6T7_ORYSA	Q5W6T7 oryza sativ
51	44	51.2	290	2	Q8YW21_ANASP	Q8YW21 anabaena sp
52	44	51.2	470	2	Q47YS4_COLP3	Q47YS4 colwellia p
53	44	51.2	492	2	Q6PGT9_BRARE	Q6PGT9 brachydanio
54	44	51.2	522	2	Q9LIW0_ORYSA	Q9LIW0 oryza sativ
55	44	51.2	1484	2	Q4Q4K3_LEIMA	Q4Q4K3 leishmania
56	44	51.2	2835	2	Q8G9Q2_LEUME	Q8G9Q2 leuconostoc
57	44	51.2	3972	2	Q9S0R8_STRAW	Q9S0R8 streptomyce
58	44	51.2	5532	2	Q9S0R4_STRAW	Q9S0R4 streptomyce
59	43.5	50.6	243	2	Q4E770_9RICK	Q4E770 wolbachia e
60	43.5	50.6	357	2	Q4ECX5_9RICK	Q4ECX5 wolbachia e
61	43	50.0	181	2	Q7NTK4_CHRVO	Q7NTK4 chromobacte
62	43	50.0	225	2	Q2WRJ4_CLOBE	Q2WRJ4 clostridium
63	43	50.0	277	2	Q3M3P1_ANAVT	Q3M3P2 anabaena va
64	43	50.0	342	1	SIA4A_CHICK	Q11200 g cmp-n-ace
65	43	50.0	375	2	Q8A7Y3_BACTN	Q8A7Y3 bacteroides
66	43	50.0	443	2	Q9A2C5_BACTN	Q9A2C5 bacteroides
67	43	50.0	479	2	Q49446_ARATH	Q49446 arabidopsis
68	43	50.0	521	2	Q5TMS4_ANOGA	Q5TMS4 anopheles g
69	43	50.0	556	2	Q7NWL5_CHRVO	Q7NWL5 chromobacte
70	43	50.0	562	2	Q5PP23_ARATH	Q5PP23 arabidopsis
71	43	50.0	581	2	Q6A3Q7_9BACT	Q6A3Q7 uncultured
72	43	50.0	592	1	HPSE2_HUMAN	Q8WWQ2 homo sapien
73	43	50.0	592	2	Q2M1H9_HUMAN	Q2M1H9 homo sapien
74	43	50.0	608	2	Q5LHS6_BACFN	Q5LHS6 bacteroides
75	43	50.0	608	2	Q4LYS8_BACFR	Q4LYS8 bacteroides
76	43	50.0	624	2	Q2USY5_ASPOR	Q2USY5 aspergillus
77	43	50.0	679	2	Q8A6U5_BACTN	Q8A6U5 bacteroides
78	43	50.0	690	2	Q2UTE2_ASPOR	Q2UTE2 aspergillus
79	43	50.0	852	2	Q5SR54_HUMAN	Q5SR54 homo sapien
80	43	50.0	857	2	Q4WN22_ASPTU	Q4WN22 aspergillus
81	43	50.0	926	2	Q4RVF3_TETNG	Q4RVF3 tetraodon n
82	43	50.0	948	2	Q9DDD3_CHICK	Q9DDD3 gallus gall
83	43	50.0	971	2	Q5UE58_HUMAN	Q5UE58 homo sapien
84	43	50.0	971	2	Q71MNO_HUMAN	Q71MNO homo sapien
85	43	50.0	971	2	Q8N4K9_HUMAN	Q8N4K9 homo sapien
86	43	50.0	978	1	CSTN1_DROME	Q9V498 drosophila
87	43	50.0	981	1	CSTN1_HUMAN	Q5SR52 homo sapien
88	43	50.0	1177	2	Q87711_PYRFU	Q87711 pyrococcus
89	43	50.0	1177	2	Q9V1R8_PYRAB	Q9V1R8 pyrococcus
90	43	50.0	1200	2	Q73340_SYNY3	Q73340 synecocyst
91	43	50.0	1291	2	Q8TYZ2_PYRFU	Q8TYZ2 pyrococcus
92	43	50.0	1348	2	Q7P2L0_ANOGA	Q7P2L0 anopheles g
93	43	50.0	1675	2	Q60EN4_ORYSA	Q60EN4 oryza sativ
94	43	50.0	383	2	Q8C3K0_MOUSE	Q8C3K0 mus musculus
95	42.5	49.4	559	2	Q3Z5D9_SHISS	Q3Z5D9 shigella so
96	42.5	49.4	559	2	Q3Z5D9_SHISS	Q3Z5D9 shigella so
97	42.5	49.4	894	2	Q2Y774_NITMU	Q2Y774 nitrospir
98	42.5	49.4	1035	2	Q7AHJ7_ECO57	Q7AHJ7 escherichia
99	42.5	49.4	1144	2	Q8X7W9_ECO57	Q8X7W9 escherichia
100	42.5	49.4	1258	2	Q80S27_MIDDL	Q80S27 middelburg
101	42	48.8	41	2	Q54QV6_DICDI	Q54QV6 dictyosteli
102	42	48.8	94	2	Q8XM53_CLOPE	Q8XM53 clostridium
103	42	48.8	129	2	Q4D9U2_TRYCR	Q4D9U2 trypanosoma
104	42	48.8	146	2	Q7R4L4_GIALA	Q7R4L4 giardia lam

105	42	48.8	180	2	Q8BJX5_MOUSE	Q8bjx5 mus musculus	178	41.5	48.3	192	2	O55932_9H1V1	O55932 human immun
106	42	48.8	181	2	Q31QA9_SYNBP7	Q31qa9 synechococc	179	41.5	48.3	192	2	O55933_9H1V1	O55933 human immun
107	42	48.8	181	2	Q5N3X7_SYNBP6	Q5n3x7 synechococc	180	41.5	48.3	192	2	O55935_9H1V1	O55935 human immun
108	42	48.8	182	2	Q357P8_9BRAD	Q357p8 bradyrhizob	181	41.5	48.3	192	2	O55936_9H1V1	O55936 human immun
109	42	48.8	201	2	Q32S45_9MOLL	Q32s45 eupyrymna sc	182	41.5	48.3	192	2	O55938_9H1V1	O55938 human immun
110	42	48.8	220	2	O5JP11_HUMAN	O5jpi1 homo sapien	183	41.5	48.3	192	2	O55941_9H1V1	O55941 human immun
111	42	48.8	221	2	Q8DWJ3_STRWD	Q8dwj3 streptococc	183	41.5	48.3	192	2	O55942_9H1V1	O55942 human immun
112	42	48.8	238	2	Q4WJ49_ASPFU	Q4wj49 aspergillus	184	41.5	48.3	192	2	O55943_9H1V1	O55943 human immun
113	42	48.8	250	2	Q7Q2C2_ANOGA	Q7q2c2 anopheles g	185	41.5	48.3	192	2	O55944_9H1V1	O55944 human immun
114	42	48.8	256	2	O5STI0_HUMAN	O5sti0 homo sapien	186	41.5	48.3	192	2	O55945_9H1V1	O55945 human immun
115	42	48.8	258	2	O5SSW3_HUMAN	O5ssw3 homo sapien	187	41.5	48.3	192	2	O55946_9H1V1	O55946 human immun
116	42	48.8	276	2	Q33IU9_9CAUD	Q33iu9 bacterioph	188	41.5	48.3	192	2	Q7ZJC5_9H1V1	Q7zjc5 human immun
117	42	48.8	325	2	Q9BEG4_BOVIN	Q9beg4 bos taurus	189	41.5	48.3	442	2	Q4QF00_DESAC	Q4qf00 desulfuromo
118	42	48.8	337	1	SI4A4_MOUSE	P54751 m cmp-n-ace	191	41.5	48.3	688	1	EFQ_APPPP	Q9zeu4 apple proli
119	42	48.8	337	1	Q3UU61_MOUSE	Q3uu61 mus musculus	191	41.5	48.3	688	2	O6C995_YARLI	Q6c995 yarrowia li
120	42	48.8	337	2	O544T4_MOUSE	O544t4 m 13 days e	192	41	47.7	66	2	O8QRN4_9CORO	O8qrn4 canine coro
121	42	48.8	340	1	MRAY_BICBP	P59436 buchnera ap	193	41	47.7	95	2	O6LXB0_METMP	O6lxb0 methanococc
122	42	48.8	340	1	SI4A4_HUMAN	Q11201 h cmp-n-ace	194	41	47.7	150	2	Q3R906_XYLFA	Q3r9q6 xylella fas
123	42	48.8	340	1	SI4A4_PANTR	Q6hb59 p cmp-n-ace	195	41	47.7	150	2	Q3R193_XYLFA	Q3r193 xylella fas
124	42	48.8	340	2	Q6H8N0_RAT	O6hb80 rattus norv	196	41	47.7	182	2	Q7U075_RHOBA	Q7uu75 rhodopirell
125	42	48.8	343	1	SI4A4_FIG	O2745 s cmp-n-ace	197	41	47.7	183	2	O89ID8_BRAJA	O89id8 bradyrhizob
126	42	48.8	343	2	Q73T71_MYCPA	O73t71 mycobacteri	198	41	47.7	216	2	Q3PWF9_NITHA	Q3pwf9 nitrobacter
127	42	48.8	357	2	Q62WX9_BACID	O62wx9 bacillus li	199	41	47.7	216	2	Q7LZU7_9CORO	Q7lzu7 transmissib
128	42	48.8	358	2	Q65LI8_BACLD	O65li8 bacillus li	200	41	47.7	218	2	Q2VYJ3_MAGSA	Q2vyj3 magnetospir
129	42	48.8	368	2	O5SSW4_HUMAN	O5ssw4 homo sapien	201	41	47.7	226	2	Q9HS16_HUMAN	Q9hs16 homo sapien
130	42	48.8	380	2	Q7MB15_VIBVY	O7mb15 vibrio vuln	202	41	47.7	254	2	Q2L2P0_BORAV	Q2l2p0 bordetella
131	42	48.8	396	2	Q8U9M4_AGR75	O8u9m4 agrobacteri	203	41	47.7	260	2	Q2NRG5_SODGL	Q2nr95 sodalis glo
132	42	48.8	419	2	O8C5S6_MOUSE	O8c5s6 mus musculus	204	41	47.7	289	2	Q2R2T1_ORYSA	Q2r2t1 oryza sativ
133	42	48.8	447	2	Q35QK5_9BRAD	Q35qk5 bradyrhizob	205	41	47.7	305	2	Q3MDL7_ANAVT	Q3mdl7 anabaena va
134	42	48.8	469	2	Q52PH1_MAIZE	Q52ph1 zea mays (m	206	41	47.7	305	2	O8YP55_ANASP	O8yp55 anabaena sp
135	42	48.8	472	1	SYG_PROAC	O6a964 propionibac	207	41	47.7	307	2	Q8UZE9_PYRFU	Q8uze9 pyrococcus
136	42	48.8	472	1	Q93XP6_MAIZE	O93xp6 zea mays (m	208	41	47.7	318	2	O4RHB6_TETNG	Q4rhb6 tetraodon n
137	42	48.8	492	2	Q7CT44_AGR75	O7ct44 agrobacteri	209	41	47.7	320	2	Q9U3M4_CAEEL	Q9u3m4 caenorhabdi
138	42	48.8	510	2	O5WD67_BACSK	O5wd67 bacillus cl	210	41	47.7	322	2	Q2R2S3_ORYSA	Q2r2s3 oryza sativ
139	42	48.8	514	2	O5SSX7_HUMAN	Q5ssx7 homo sapien	211	41	47.7	331	2	Q3R0X1_XYLFA	Q3r0x1 xylella fas
140	42	48.8	516	2	O5STI1_HUMAN	O5sti1 homo sapien	212	41	47.7	333	2	Q3R5B9_XYLFA	Q3r5b9 xylella fas
141	42	48.8	533	2	Q73F21_BACC1	Q73f21 bacillus ce	213	41	47.7	353	2	Q9PB56_XYLFA	Q9pb56 xylella fas
142	42	48.8	536	2	Q3GZ29_9ACTO	Q3gz29 nocardioid	214	41	47.7	367	2	O2S9G3_9GAMW	Q2s9g3 habella che
143	42	48.8	554	1	ARD1_RAT	P36407 rattus norv	215	41	47.7	374	2	O5NXV2_AZOSE	O5nxv2 azoarcus sp
144	42	48.8	574	1	ARD1_HUMAN	P36406 homo sapien	216	41	47.7	388	2	Q3WQ7_9ACTO	Q3wiq7 frankia sp
145	42	48.8	574	1	ARD1_MOUSE	Q8bgx0 mus musculus	217	41	47.7	390	2	Q347I8_RHOFA	Q347i8 rhodospseudo
146	42	48.8	588	2	Q6DDN3_XENLA	Q6ddn3 xenopus lae	218	41	47.7	411	2	Q35R75_9BRAD	Q35r75 bradyrhizob
147	42	48.8	625	2	Q4S258_TETNG	Q4s258 tetraodon n	219	41	47.7	411	2	Q7G780_ORYSA	Q7g780 oryza sativ
148	42	48.8	660	2	Q4W8R4_HORVU	Q4w8r4 hordeum vul	220	41	47.7	428	2	O24241_ORYSA	O24241 porcine res
149	42	48.8	664	1	RN139_HUMAN	O8wul7 homo sapien	221	41	47.7	428	2	Q02422_9CORO	Q02422 porcine res
150	42	48.8	664	2	Q5RBT7_PONPY	O5rbt7 pongo pygma	222	41	47.7	428	2	Q02423_9CORO	Q02423 porcine res
151	42	48.8	668	1	RN139_MOUSE	Q7tmv1 mus musculus	223	41	47.7	428	2	Q02424_9CORO	Q02424 porcine res
152	42	48.8	688	2	Q5Z168_CHICK	O5z168 gallus gall	224	41	47.7	456	2	Q4G033_RAT	Q4g033 rattus norv
153	42	48.8	767	2	Q9HID7_THEAC	O9hid7 thermoplasm	225	41	47.7	463	2	O59811_SCHPO	O59811 schizosacch
154	42	48.8	855	2	Q4SKR4_TETNG	O4skr4 tetraodon n	226	41	47.7	473	2	Q9JVP7_NEIMA	Q9jvp7 neisseria m
155	42	48.8	886	2	Q43KS2_SOLUS	Q43ks2 solibacter	227	41	47.7	473	2	Q9K0M2_NEIME	Q9k0m2 neisseria m
156	42	48.8	956	1	CSTN3_HUMAN	Q9bqt9 homo sapien	228	41	47.7	476	2	Q8MT72_DROME	Q8mt72 drosophila
157	42	48.8	956	1	CSTN3_MOUSE	Q99j17 mus musculus	229	41	47.7	490	1	SYG_BIFLO	Q9g7w9 bifidobacte
158	42	48.8	956	2	Q2T9J5_HUMAN	Q2t9j5 homo sapien	230	41	47.7	490	2	O5EMH2_9CORO	Q5emh2 canine coro
159	42	48.8	956	2	Q5R9Q9_PONPY	O5r9q9 pongo pygma	231	41	47.7	490	2	O5G1S4_9CORO	Q5g1s4 canine coro
160	42	48.8	956	2	O544R0_MOUSE	O544r0 mus musculus	232	41	47.7	497	2	O5F6Y2_NEIG1	O5f6y2 neisseria g
161	42	48.8	957	1	CSTN3_RAT	O8r553 rattus norv	233	41	47.7	504	2	Q37W36_SPHAR	Q37w36 novosphingo
162	42	48.8	968	2	Q5UB57_HUMAN	Q5ue57 homo sapien	234	41	47.7	507	2	Q74MZ6_NANEQ	Q74mz6 nanoarchaeu
163	42	48.8	1100	2	Q61BN8_CAEER	Q61bn8 caenorhabdi	235	41	47.7	511	1	GPMT_STRCO	Q91214 streptomyce
164	42	48.8	1155	2	Q6TKT8_ECOLI	O6tkt8 escherichia	236	41	47.7	542	2	O5FQW3_GLUOX	Q5fqw3 gluconobact
165	42	48.8	1195	2	Q9PUM2_XENLA	O9pum2 xenopus lae	237	41	47.7	578	2	O6ZUX8_HUMAN	O6zux8 homo sapien
166	42	48.8	1501	2	Q27720_PLAFA	Q27720 plasmodium	238	41	47.7	609	2	Q6ZU58_HUMAN	Q6zu58 homo sapien
167	42	48.8	1555	2	Q8U514_PLAF7	Q8u514 plasmodium	239	41	47.7	613	2	O4IB63_GIBZE	Q4ib63 gibberella
168	42	48.8	1748	2	Q9U421_PLAFA	Q9u421 plasmodium	240	41	47.7	621	2	Q4SPT3_TETNG	Q4sft3 tetraodon n
169	42	48.8	1748	2	Q4Q3F1_LEBETA	Q4q3f1 leishmania	241	41	47.7	624	2	Q74SQ4_MYCPA	Q74sq4 mycobacteri
170	41.5	48.3	117	2	Q8QRX7_9BETA	Q8qrx7 pongine her	242	41	47.7	639	2	Q9WJ39_DROME	Q9wj39 drosophila
171	41.5	48.3	192	2	O55926_9H1V1	O55926 human immun	243	41	47.7	668	2	Q4P6C8_UBTMA	Q4p6c8 uetillago ma
172	41.5	48.3	192	2	O55926_9H1V1	O55926 human immun	244	41	47.7	687	2	Q9DUT2_9BTOM	Q9dut2 pelargonium
173	41.5	48.3	192	2	O55927_9H1V1	O55927 human immun	245	41	47.7	693	2	O5W7F3_BOMMO	O5w7f3 bombyx mori
174	41.5	48.3	192	2	O55928_9H1V1	O55928 human immun	246	41	47.7	753	2	Q628L0_CAEER	Q628l0 caenorhabdi
175	41.5	48.3	192	2	O55929_9H1V1	O55929 human immun	247	41	47.7	777	2	Q9QRQ1_9CORO	Q9qrq1 canine coro
176	41.5	48.3	192	2	O55930_9H1V1	O55930 human immun	248	41	47.7	784	2	O65I20_BACILD	O65i20 bacillus li
177	41.5	48.3	192	2	O55931_9H1V1	O55931 human immun	249	41	47.7	789	2	Q8QRR7_9CORO	Q8qrr7 transmissib

251	41	47.7	789	2	QBQR8_9COCO	QBqrr8	transmissib	324	40	46.5	121	2	Q926M3_LISIN	Q926m3	listeria in
252	41	47.7	789	2	QBQR9_9COCO	QBqrr9	transmissib	325	40	46.5	127	1	ILBP_RABIT	P50119	oryctolagus
253	41	47.7	789	2	QBQR50_9COCO	QBqrs0	transmissib	326	40	46.5	132	2	O22229_CABEL	O62229	caenorhabdi
254	41	47.7	789	2	QBQR53_9COCO	QBqrs3	transmissib	327	40	46.5	137	2	Q9ACRI_STRCO	Q9acri	streptomyc
255	41	47.7	790	2	Q5LIY7_GEOKA	Q5liy7	geobacillus	328	40	46.5	150	2	Q5KUW8_GROKA	Q8q5u8	geobacillus
256	41	47.7	790	2	QBQR51_9COCO	QBqrs1	transmissib	329	40	46.5	155	2	Q8HG55_PSEPK	Q8hg55	pseudomonas
257	41	47.7	790	2	QBQR52_9COCO	QBqrs2	transmissib	330	40	46.5	182	2	Q5L246_GROKA	Q8l246	geobacillus
258	41	47.7	793	2	QBQRQ2_9COCO	QBqrq2	canine coro	331	40	46.5	196	2	Q4ZMJ5_PSEU2	Q4zmj5	pseudomonas
259	41	47.7	808	2	QBQRQ2_9COCO	QBqrq2	canine coro	332	40	46.5	196	2	Q88A22_PSESJ	Q88a22	pseudomonas
260	41	47.7	813	2	QBQRQ4_9COCO	QBqrq4	canine coro	333	40	46.5	246	2	Q2UBJ1_ASPOR	Q2ubj1	aspergillus
261	41	47.7	815	2	QBQRQ3_9COCO	QBqrq3	canine coro	334	40	46.5	252	2	Q63LW5_BURPS	Q63lw5	burkholderi
262	41	47.7	883	2	Q55K31_CRYNE	Q55k31	cryptococcu	335	40	46.5	256	2	Q44XV6_9BURK	Q44xv6	burkholderi
263	41	47.7	883	2	Q5K9G8_CRYNE	Q5k9g8	cryptococcu	336	40	46.5	256	2	Q41V13_9BURK	Q41v13	burkholderi
264	41	47.7	941	2	Q4PBR4_USTMA	Q4pbr4	ustilago ma	337	40	46.5	256	2	Q3JFK8_BURP1	Q3jfk8	burkholderi
265	41	47.7	1064	2	Q4J7S9_SULAC	Q4j7s9	ustilobus	338	40	46.5	256	2	Q2T5S3_BURTH	Q2t5s3	burkholderi
266	41	47.7	1127	2	Q5CS41_CRYPV	Q5cs41	cryptospori	339	40	46.5	256	2	Q62BH7_BURMA	Q62bh7	burkholderi
267	41	47.7	1265	1	SPIKE_CVPR8	P27655	porcine res	340	40	46.5	268	2	Q935I7_SALTI	Q935i7	salmonella
268	41	47.7	1225	1	SPIKE_CVPRM	P24413	porcine res	341	40	46.5	274	2	Q8TIG8_METAC	Q8tig8	methanosa
269	41	47.7	1225	2	Q84852_9COCO	Q84852	porcine res	342	40	46.5	282	2	Q3Q562_9GNMM	Q3q562	shewanella
270	41	47.7	1263	2	Q7Z485_HUMAN	Q7z485	homo sapien	343	40	46.5	282	2	Q3Q5G8_9BRAD	Q3q5g8	bradyrhizob
271	41	47.7	1310	1	AT8B3_HUMAN	Q60423	homo sapien	344	40	46.5	284	1	TRUA_SYNEL	Q8cvm6	synecococc
272	41	47.7	1393	2	Q338U3_ORYSA	Q338u3	oryza sativ	345	40	46.5	291	2	Q3WML3_9RHIZ	Q3wwl3	mesorhizobi
273	41	47.7	1429	2	Q5CLH5_CRYHO	Q5clh5	cryptospori	346	40	46.5	324	2	Q5LHS5_BACFN	Q5lhs5	bacteroides
274	41	47.7	1429	2	Q5CV19_CRYPV	Q5cv19	cryptospori	347	40	46.5	324	2	Q64YS7_BACFR	Q64ys7	bacteroides
275	41	47.7	1447	1	SPIKE_CVPPU	P07946	porcine tra	348	40	46.5	325	2	Q3GHJ3_CHLVI	Q3ghj3	prosthecoc
276	41	47.7	1447	1	SPIKE_CVPR7	Q19777	porcine tra	349	40	46.5	330	2	Q2KBS5_FUGRU	Q2kbs5	fugu rubrip
277	41	47.7	1447	1	QBQRQ0_9COCO	QBqrq0	transmissib	350	40	46.5	332	2	Q70DS6_XENTR	Q70ds6	xenopus tro
278	41	47.7	1447	1	Q9IWO4_9COCO	P18450	porcine tra	351	40	46.5	332	2	Q70DS7_XENLA	Q70ds7	xenopus lae
279	41	47.7	1449	1	SPIKE_CVPPS	P18450	porcine tra	352	40	46.5	334	2	Q3AWN9_SYNS9	Q3awn9	synecococc
280	41	47.7	1449	1	SPIKE_CVPMI	P33470	porcine tra	353	40	46.5	337	2	Q2V3S5_ARATH	Q2v3s5	arabidopsia
281	41	47.7	1449	2	Q4ZJU7_9COCO	Q4zju7	transmissib	354	40	46.5	337	2	Q7MYQ1_PHOLL	Q7myq1	photorhabd
282	41	47.7	1449	2	QBQRQ5_9COCO	QBqrq5	transmissib	355	40	46.5	338	2	Q5U496_XENLA	Q5u496	xenopus lae
283	41	47.7	1449	2	Q7T428_9COCO	Q7t428	transmissib	356	40	46.5	344	2	Q8A533_BACTN	Q8a533	bacteroides
284	41	47.7	1449	2	Q84853_9COCO	Q84853	porcine res	357	40	46.5	350	2	Q3SLH7_9BRAD	Q3slh7	bradyrhizob
285	41	47.7	1449	2	Q88510_9COCO	Q88510	transmissib	358	40	46.5	358	2	Q638V1_BACC2	Q638v1	bacillus ce
286	41	47.7	1449	2	Q9D22_9COCO	Q9d22	transmissib	359	40	46.5	374	2	Q9N6B1_LEIMA	Q9n6b1	leishmania
287	41	47.7	1449	2	Q9YRA7_9COCO	Q9yra7	transmissib	360	40	46.5	375	2	Q4HS04_CAMUP	Q4hs04	campylobact
288	41	47.7	1449	2	Q2LJ83_9COCO	Q2lj83	transmissib	361	40	46.5	380	2	Q5ND66_ORILA	Q5nd66	oryzias lac
289	41	47.7	1451	1	SPIKE_CVCAI	P36300	canine ente	362	40	46.5	384	2	Q470X9_RALEU	Q470x9	raistonla e
290	41	47.7	1452	1	SPIKE_FTPV	P10033	feline infe	363	40	46.5	385	2	Q3RQY0_RALME	Q3rqy0	raistonla m
291	41	47.7	1452	2	Q4U5G0_9COCO	Q4u5g0	feline coro	364	40	46.5	385	2	Q937N9_RALEU	Q937n9	raistonla e
292	41	47.7	1452	2	Q52PA3_9COCO	Q52pa3	feline infe	365	40	46.5	385	2	Q5QYB3_IDILO	Q5qyb3	idionarina
293	41	47.7	1453	1	SPIKE_CVCAE	Q65984	canine ente	366	40	46.5	387	2	Q8XTI7_RALSO	Q8xti7	raistonla s
294	41	47.7	1453	2	Q6T522_9COCO	Q6t522	canine coro	367	40	46.5	388	2	Q7N1D5_PHOLL	Q7n1d5	photorhabd
295	41	47.7	1453	2	Q6TPL2_9COCO	Q6tpl2	canine coro	368	40	46.5	389	2	Q82568_ARATH	Q82568	arabidopsia
296	41	47.7	1453	2	Q7T6T3_9COCO	Q7t6t3	canine coro	369	40	46.5	390	2	Q3JHQ9_BURP1	Q3jhg9	burkholderi
297	41	47.7	1453	2	Q5RZ64_9COCO	Q5rz64	canine coro	370	40	46.5	390	2	Q2T373_BURTH	Q2t373	burkholderi
298	41	47.7	1454	2	Q66928_9COCO	Q66928	feline coro	371	40	46.5	390	2	Q62A68_BURMA	Q62a68	burkholderi
299	41	47.7	1454	2	Q6SR88_9COCO	Q6sr88	feline infe	372	40	46.5	390	2	Q63NU2_BURPS	Q63nu2	burkholderi
300	41	47.7	1472	2	Q87B49_XYLFT	Q87b49	xylella fas	373	40	46.5	393	2	Q46YU2_RALEU	Q46yu2	raistonla e
301	41	47.7	1521	2	Q3R2H6_XYLFA	Q3r2h6	xylella fas	374	40	46.5	398	2	Q2Z1A8_RALEU	Q2z1a8	raistonla e
302	41	47.7	1521	2	Q3RA73_XYLFA	Q3ra73	xylella fas	375	40	46.5	400	2	Q2L071_BORAV	Q2l071	bordetella
303	41	47.7	1521	2	Q3RUH8_XYLFA	Q3ruh8	xylella fas	376	40	46.5	400	2	Q7VW96_BORPE	Q7vw96	bordetella
304	41	47.7	2157	2	Q9AYB5_ORYSA	Q9ayb5	oryza sativ	377	40	46.5	400	2	Q7WSQ6_BORPA	Q7wsq6	bordetella
305	41	47.7	2174	2	Q2QGT7_ORYSA	Q2qgt7	oryza sativ	378	40	46.5	400	2	Q7WD92_BORBR	Q7wd92	bordetella
306	41	47.7	4268	2	Q4QFM9_LEIMA	Q4qfm9	leishmania	379	40	46.5	410	2	Q3F1E2_BACTI	Q3fie2	bacillus th
307	41	47.7	6239	2	Q9S0R7_STRAW	Q9s0r7	streptomyc	380	40	46.5	422	2	Q5MIW4_AEDAL	Q5miw4	aedes albop
308	40.5	47.1	192	2	Q55940_9HIV1	Q55940	human immun	381	40	46.5	435	2	Q4EUT3_LISMO	Q4eut3	listeria mo
309	40.5	47.1	192	2	Q55946_9HIV1	Q55946	human immun	382	40	46.5	462	2	Q72HW2_THET2	Q72hw2	thermus the
310	40.5	47.1	192	2	Q55969_9HIV1	Q55969	human immun	383	40	46.5	462	2	Q6IMZ0_RAT	Q6imz0	rattus norv
311	40.5	47.1	250	1	Y2406_ANAP	Q44510	anabaena sp	384	40	46.5	462	2	Q923M2_RAT	Q923m2	rattus norv
312	40.5	47.1	271	2	Q3QH95_9GAMM	Q3qh95	shewanella	385	40	46.5	471	2	Q5DB10_SCHJA	Q5db10	schistosoma
313	40.5	47.1	308	2	Q8FCZ9_ECOL6	Q8fcz9	escherichia	386	40	46.5	478	2	Q49QX8_SCHMA	Q49qx8	schistosoma
314	40.5	47.1	312	2	Q7XEE3_ORYSA	Q7xee3	oryza sativ	387	40	46.5	485	2	Q5DHS0_SCHJA	Q5dhs0	schistosoma
315	40.5	47.1	319	2	Q2V079_CABEL	Q2v079	caenorhabdi	388	40	46.5	508	2	Q8N9M5_HUMAN	Q8n9m5	homo sapien
316	40.5	47.1	323	2	Q95X80_CABEL	Q95x80	caenorhabdi	389	40	46.5	529	2	Q61TI9_CAEBR	Q61ti9	caenorhabdi
317	40.5	47.1	324	2	Q17172_CABEL	O17172	caenorhabdi	390	40	46.5	534	2	Q93YH2_LYCER	Q93yh2	lycopersico
318	40.5	47.1	413	2	Q2YCH3_NITMU	Q2ych3	nitrosospi	391	40	46.5	541	2	Q3HTK2_CHLRE	Q3htk2	chlamydomon
319	40	46.5	74	2	Q9K0F7_NEIMB	Q9k0f7	neisseria m	392	40	46.5	563	2	Q4SCB8_TETNG	Q4scb8	tetradodon n
320	40	46.5	81	2	Q5FA07_NEIGI	Q5fa07	neisseria g	393	40	46.5	572	2	Q8Y842_LISMO	Q8y842	listeria mo
321	40	46.5	81	2	Q9JVF9_NEINWA	Q9jvf9	neisseria m	394	40	46.5	578	2	Q47BW9_DSCAR	Q47bw9	dechloromon
322	40	46.5	105	2	Q3U595_MOUSE	Q3u595	mus musculus	395	40	46.5	592	2	Q7SY37_BRARE	Q7sy37	brachydanio
323	40	46.5	107	2	Q3WAP6_9ACTO	Q3wap6	frankia sp.	396	40	46.5					

397	40	46.5	626	2	Q5BCZ0_EMENI	Q5bcz0 aspergillus	470	39	45.3	240	2	Q9GTN2_DROSI	Q9gtn2 drosophila
398	40	46.5	634	2	Q2WIM4_CLOBE	Q2wim4 clostridium	471	39	45.3	240	2	Q9V402_DROME	Q9v402 drosophila
399	40	46.5	637	2	Q6SWU5_ORYSA	Q6swu5 oryza sativ	472	39	45.3	241	2	Q3FSY8_RHOFER	Q3fsy8 rhodospira
400	40	46.5	658	2	Q8UJ11_LACPL	Q8uj11 lactobacilli	473	39	45.3	256	2	Q5ZNV2_VIVIRU	Q5znv2 cotesia con
401	40	46.5	670	1	ARPP_ARATH	Q93vr9 arabidopsis	474	39	45.3	261	1	CALB1_CHICK	P04354 gallus gall
402	40	46.5	677	2	Q7PP32_ANOGA	Q7pp32 anopheles g	475	39	45.3	262	2	Q8PX61_METWA	Q8px61 methanosarc
403	40	46.5	686	2	Q3VAT3_9SPHN	Q3vat3 sphingopyx1	476	39	45.3	262	2	Q8TUK3_METAC	Q8tuk3 methanosarc
404	40	46.5	696	2	Q2SG98_9GAMM	Q2sg98 shigella che	477	39	45.3	263	2	Q5TJQ21_HUMAN	Q5tjq21 homo sapien
405	40	46.5	720	2	Q16270_CAEBL	Q16270 caenorhabdi	478	39	45.3	263	2	Q38UP2_LACSS	Q38up2 lactobacill
406	40	46.5	723	2	Q3GUH0_9ACTO	Q3guh0 nocardioid	479	39	45.3	285	2	Q9AIT5_VIBCH	Q9ait5 vibrio chol
407	40	46.5	723	2	Q97GM4_CLOAB	Q97gm4 clostridium	480	39	45.3	288	2	Q7ZVN0_BRARE	Q7zvn0 brachydanio
408	40	46.5	724	2	Q9AD25_STRCO	Q9ad25 streptomyce	481	39	45.3	292	2	Q81DC5_BACCR	Q81dc5 bacillus ce
409	40	46.5	841	2	Q316Q4_DESDG	Q316q4 desulfovibr	482	39	45.3	294	1	WRK70_ARATH	Q91y00 arabidopsis
410	40	46.5	889	2	Q6PYI2_OSTTA	Q6pyi2 ostreococcu	483	39	45.3	295	2	Q4WLN0_ASPEU	Q4wln0 aspergillus
411	40	46.5	971	1	QYIA_CAEEL	Q9281 caenorhabdi	484	39	45.3	299	2	Q4ZP35_PSEU2	Q4zrp35 pseudomonas
412	40	46.5	1146	1	YHC3_YEAST	P38742 saccharomyc	485	39	45.3	305	2	Q62ZG8_BACLD	Q62zg8 bacillus li
413	40	46.5	1162	2	Q7PMJ6_ANOGA	P38742 anopheles g	486	39	45.3	310	2	Q50J74_BRARE	Q50j74 brachydanio
414	40	46.5	1189	2	Q5JJA2_PYRKO	Q5jjj2 pyrococcus	487	39	45.3	313	2	Q2NR31_SODGL	Q2nr31 sodalis glo
415	40	46.5	1472	2	Q9PG24_XYLFA	Q9pg24 xyella fas	488	39	45.3	319	2	Q5NX35_AZOSE	Q5nx35 azoarcus sp
416	40	46.5	1614	2	Q5TRC9_ANOGA	Q5trc9 anopheles g	489	39	45.3	319	2	Q6SP29_BACLD	Q6sp29 bacillus li
417	40	46.5	1675	2	Q582S2_9TRYP	Q582s2 trypanosoma	490	39	45.3	320	2	Q9LRQ2_ARATH	Q9lrq2 arabidopsis
418	39.5	45.9	192	2	Q4JGH0_9HIV1	Q4jgh0 human immun	491	39	45.3	321	2	Q66DG1_YERPS	Q66dg1 yersinia ps
419	39.5	45.9	192	2	Q4U541_9HIV1	Q4u541 human immun	492	39	45.3	321	2	Q8ZDH6_YERPE	Q8zdh6 yersinia pe
420	39.5	45.9	257	2	Q953X6_9BILA	Q953x6 terebratali	493	39	45.3	330	2	Q5TIN8_BRARE	Q5tin8 brachydanio
421	39.5	45.9	310	2	Q72MN6_LEPTC	Q72mn6 leptospira	494	39	45.3	331	2	Q6KB56_TETNG	Q6kb56 tetraodon n
422	39.5	45.9	324	2	Q8EZA3_LEPIN	Q8eza3 leptospira	495	39	45.3	337	2	Q20269_CAEBL	Q20269 caenorhabdi
423	39.5	45.9	310	2	Q47B91_DECAR	Q47b91 dechloromon	496	39	45.3	339	2	Q7UXZ2_RHOBA	Q7uxz2 rhodopirell
424	39.5	45.9	411	2	Q8QOM3_METWA	Q8qom3 methanosarc	497	39	45.3	343	1	FUT2_HYLLA	Q9ctc7 hylobates l
425	39.5	45.9	611	2	Q5RIF1_BRARE	Q5rif1 brachydanio	498	39	45.3	346	2	Q3E7K9_ARATH	Q3e7k9 arabidopsis
426	39.5	45.9	611	2	Q6NUX4_BRACHD	Q6nux4 brachydanio	499	39	45.3	348	2	Q4QAL8_LEITWA	Q4qal8 leishmania
427	39.5	45.9	612	2	Q5ZJY6_CHICK	Q5zjj6 gallus gall	500	39	45.3	358	2	Q8U2T3_PYRFU	Q8u2t3 pyrococcus
428	39.5	45.9	630	2	Q4T8V9_TETNG	Q4t8v9 tetraodon n	501	39	45.3	365	2	Q59565_PYRHO	Q59565 pyrococcus
429	39.5	45.9	697	2	Q6CQK9_KLUJIA	Q6cqk9 kluyveromyc	502	39	45.3	365	2	Q9V213_PYRAB	Q9v213 pyrococcus
430	39.5	45.9	1867	2	Q2U2A2_ASPOR	Q2u2a2 aspergillus	503	39	45.3	370	2	Q2LQX0_9DELT	Q2lqx0 syntrophus
431	39	45.3	59	2	Q349J1_RHOPA	Q349j1 rhodopsendo	504	39	45.3	371	2	Q29919_ARCFU	Q29919 archaeoglob
432	39	45.3	59	2	Q34BD4_RHOPA	Q34bd4 rhodopsendo	505	39	45.3	371	2	Q59Y10_HUMAN	Q59y10 homo sapien
433	39	45.3	59	2	Q36VP4_RHOPA	Q36vp4 rhodopsendo	506	39	45.3	373	2	Q8AB77_BACTN	Q8ab77 bacteroides
434	39	45.3	59	2	Q36VP6_RHOPA	Q36vp6 rhodopsendo	507	39	45.3	377	2	Q4ZVV0_PSEU2	Q4zvv0 pseudomonas
435	39	45.3	59	2	Q37BH8_RHOPA	Q37bh8 rhodopsendo	508	39	45.3	389	2	Q04178_BRACM	Q04178 brassica ca
436	39	45.3	59	2	Q37B67_RHOPA	Q37b67 rhodopsendo	509	39	45.3	389	2	Q82567_ARATH	Q82567 arabidopsis
437	39	45.3	59	2	Q37FH2_RHOPA	Q37fh2 rhodopsendo	510	39	45.3	389	2	Q39810_SOYBN	Q39810 glycine max
438	39	45.3	59	2	Q218T5_RHOPA	Q218t5 rhodopsendo	511	39	45.3	389	2	Q9SP57_BRACM	Q9sp57 brassica ca
439	39	45.3	59	2	Q21W45_RHOPA	Q21w45 rhodopsendo	512	39	45.3	392	2	Q8WIE9_BRARP	Q8wie9 brassica ra
440	39	45.3	64	2	Q3EVA0_BACTI	Q3eva0 bacillus th	513	39	45.3	392	2	Q81A41_BACCR	Q81a41 bacillus ce
441	39	45.3	90	2	Q4R9Y9_TETNG	Q4r9y9 tetraodon n	514	39	45.3	400	2	Q7V7D6_PROMM	Q7v7d6 prochloroco
442	39	45.3	99	2	Q74LF6_LACJO	Q74lf6 lactobacilli	515	39	45.3	402	2	Q4WCA8_ASPEU	Q4wca8 aspergillus
443	39	45.3	106	2	Q2W8L0_MAGSA	Q2w8l0 magnetospir	516	39	45.3	408	2	Q5BDP8_EMENI	Q5bdp8 aspergillus
444	39	45.3	108	2	Q9KEC8_BACHD	Q9kec8 bacillus ha	517	39	45.3	410	2	Q7CMD2_BACAN	Q7cmd2 bacillus an
445	39	45.3	133	2	Q2Z903_9GAMM	Q2z903 shewanella	518	39	45.3	410	2	Q9X367_BACILL	Q9x367 bacillus an
446	39	45.3	135	2	Q05499_BACSU	Q05499 bacillus su	519	39	45.3	416	2	Q74NL9_BACCI	Q74nl9 bacillus bo
447	39	45.3	147	2	Q17986_CAEBL	Q17986 caenorhabdi	520	39	45.3	417	2	Q31V01_SHIBS	Q31v01 shigella bo
448	39	45.3	159	2	Q7R5F3_GIALA	Q7r5f3 giardia lam	521	39	45.3	417	2	Q3YVY6_SHISS	Q3yvvy6 shigella so
449	39	45.3	161	2	Q3U6U3_MOUSE	Q3u6u3 mus musculu	522	39	45.3	417	2	Q8KMW4_ECOLI	Q8kmw4 escherichia
450	39	45.3	170	2	Q9VX72_DROME	Q9vx72 drosophila	523	39	45.3	417	2	Q9ZIT0_ECOLI	Q9zit0 escherichia
451	39	45.3	172	2	Q4MGF7_BACCE	Q4mgf7 bacillus ce	524	39	45.3	421	2	Q8FC97_RHOLA	Q8fc97 rhodospirell
452	39	45.3	173	2	Q8U2X6_PYRFU	Q8u2x6 pyrococcus	525	39	45.3	421	2	Q7UUD0_RHOPA	Q7uud0 rhodopirell
453	39	45.3	175	2	Q9VD84_DROME	Q9vd84 drosophila	526	39	45.3	422	2	Q5Z513_ORYSA	Q5z513 oryza sativ
454	39	45.3	179	2	Q5UQ20_HUMAN	Q5jq20 homo sapien	527	39	45.3	426	2	Q8A6V4_BACTN	Q8a6v4 bacteroides
455	39	45.3	183	2	Q55AD9_DICDI	Q55ad9 dictyosteli	528	39	45.3	427	2	Q5TQ40_ANOGA	Q5tq40 anopheles g
456	39	45.3	186	1	ARL6_MOUSE	Q88848 mus musculu	529	39	45.3	433	2	Q2SF93_9GAMM	Q2sf93 hanelia che
457	39	45.3	186	2	Q73QC5_TREDE	Q73qc5 treponema d	530	39	45.3	451	2	Q9LRQ7_ARATH	Q9lrq7 arabidopsis
458	39	45.3	186	2	Q3TUM2_MOUSE	Q3tum2 mus musculu	531	39	45.3	452	2	Q9FNP9_ARATH	Q9fnp9 arabidopsis
459	39	45.3	186	2	Q8VCV3_MOUSE	Q8vcv3 mus musculu	532	39	45.3	455	2	Q5Z514_ORYSA	Q5z514 oryza sativ
460	39	45.3	187	2	Q684F5_VIVIRU	Q684f5 sulfolobus	533	39	45.3	455	2	Q93XP5_ORYSA	Q93xp5 oryza sativ
461	39	45.3	193	2	Q3TY77_MOUSE	Q3ty77 mus musculu	534	39	45.3	458	2	Q82XT9_NITEU	Q82xt9 nitrosomona
462	39	45.3	197	2	Q5QSC8_9HIV1	Q5qsc8 human immun	535	39	45.3	471	2	Q3GYW7_9ACTO	Q3gyw7 nocardioid
463	39	45.3	199	2	Q84D96_9BACT	Q84d96 uncultured	536	39	45.3	477	2	Q5KQZ4_KLEPN	Q5kqz4 klebsiella
464	39	45.3	209	2	Q2L039_BORAV	Q2l039 bordetella	537	39	45.3	477	2	Q5MQC2_ECOLI	Q5mqc2 escherichia
465	39	45.3	213	2	Q2UTX6_ASPOR	Q2utx6 aspergillus	538	39	45.3	483	2	Q4XQP1_RALSO	Q4xqp1 raietonia s
466	39	45.3	218	2	Q518X8_PLAFA	Q5i8x8 plasmodium	539	39	45.3	493	2	Q43D94_CHLBO	Q43d94 chlorobium
467	39	45.3	218	2	Q8IINI_PLAF7	Q8iini plasmodium	540	39	45.3	504	2	Q5XL56_KLEPN	Q5xl56 klebsiella
468	39	45.3	230	2	Q6UCP8_9PROT	Q6ucp8 uncultured	541	39	45.3	515	2	Q9SIU7_ARATH	Q9siu7 arabidopsis
469	39	45.3	236	2	Q4PB40_UMTWA	Q4pe40 ustilago ma	542	39	45.3	523	2	Q8Z5B6_STRAW	Q8z5b6 streptomyce

543	39	45.3	534	1	TYRPL_AMBME	P55027 ambystoma m	616	39	45.3	1003	2	Q6VLI9_HUMAN	Q6vli9 homo sapien
544	39	45.3	547	2	Q6PBV6_ACIAD	Q6fbv6 acinetobact	617	39	45.3	1035	2	Q69ZV9_MOUSE	Q69zv9 mus musculu
545	39	45.3	551	1	ASLA_ECOLI	P25549 escherichia	618	39	45.3	1050	2	Q50XRS_ENTHI	Q50xrs encamoeba h
546	39	45.3	551	2	Q329X6_SHIDS	Q229x6 shigella dy	619	39	45.3	1072	2	Q4SL50_TETNG	Q4sl50 tetraodon n
547	39	45.3	551	2	Q2M8A6_ECOLI	Q2m8a6 escherichia	620	39	45.3	1085	2	Q68D12_HUMAN	Q68d12 homo sapien
548	39	45.3	551	2	Q8XAO0_ECO57	Q8xao0 escherichia	621	39	45.3	1137	1	KCMW1_CHICK	Q8ay88 g calcium-a
549	39	45.3	554	2	Q72MY0_LEPIC	Q72my0 leptospira	622	39	45.3	1144	2	Q9W7J2_TRASC	Q9w7j2 trachemys s
550	39	45.3	554	2	Q8EZI3_LEPIN	Q8ezi3 leptospira	623	39	45.3	1151	1	KCMW1_MACHU	Q18867 m calcium-a
551	39	45.3	559	2	Q500V2_ARATH	Q500v2 arabidopsis	624	39	45.3	1152	1	KCMW1_PIG	Q18866 s calcium-a
552	39	45.3	562	2	Q7QCQ5_ANOAG	Q7qcg5 anopheles g	625	39	45.3	1165	2	Q6DIL4_ERWCT	Q6dil4 erwinia car
553	39	45.3	583	2	Q9VHN8_DROME	Q9vhn8 drosophila	626	39	45.3	1166	1	KCMW1_BOVIN	Q28204 b calcium-a
554	39	45.3	590	2	Q59FH2_HUMAN	Q59fh2 homo sapien	627	39	45.3	1166	2	Q5SOR7_HUMAN	Q5sor7 homo sapien
555	39	45.3	591	1	Y1280_MYCTU	P66771 mycobacteri	628	39	45.3	1166	2	Q73728_TRASC	Q73728 trachemys s
556	39	45.3	591	1	Y1311_MYCBO	P66772 mycobacteri	629	39	45.3	1167	2	Q5SQS1_HUMAN	Q5sq81 homo sapien
557	39	45.3	595	2	Q59VF9_CANAL	Q59vf9 candida alb	630	39	45.3	1171	2	Q5SQRC_HUMAN	Q5sq86 homo sapien
558	39	45.3	605	2	Q2ZXZ1_STRSU	Q2zxz1 streptococc	631	39	45.3	1173	2	Q73729_TRASC	Q73729 trachemys s
559	39	45.3	607	2	Q4ZLQ6_PSEU2	Q4zlg6 pseudomonas	632	39	45.3	1177	2	Q5SQRS_HUMAN	Q5sq89 homo sapien
560	39	45.3	618	2	Q9HIW0_THEAC	Q9hiw0 thermoplasm	633	39	45.3	1178	2	Q5JQ23_HUMAN	Q5jq23 homo sapien
561	39	45.3	621	2	Q4QH29_LEIMA	Q4qhz9 leishmania	634	39	45.3	1179	1	KCMW1_RABIT	Q9bg98 o calcium-a
562	39	45.3	624	2	Q65VV9_NANSM	Q65vv9 manheimia	635	39	45.3	1181	2	Q5SQR8_HUMAN	Q5sq88 homo sapien
563	39	45.3	626	2	Q9VKU1_DROME	Q9vku1 drosophila	636	39	45.3	1181	2	Q9KN45_VIBCH	Q9kn45 vibrio chol
564	39	45.3	627	2	Q3AFG3_CARHZ	Q3afg3 carboxydoth	637	39	45.3	1192	2	Q3IT35_NATPD	Q3it35 natronomona
565	39	45.3	637	2	Q8BIG5_MOUSE	Q8big5 mus musculu	638	39	45.3	1196	1	KCMW1_XENLA	Q90zc7 x calcium-a
566	39	45.3	652	2	Q4NLW4_WMICC	Q4nlw4 arthrobacte	639	39	45.3	1196	2	Q7NQR1_CHRVO	Q7nqr1 chromobacte
567	39	45.3	656	2	Q8CLP5_YERPE	Q8clp5 yersinia pe	640	39	45.3	1200	2	Q73731_TRASC	Q73731 trachemys s
568	39	45.3	662	2	Q6MU55_MYCWS	Q6mu55 mycoplasma	641	39	45.3	1203	1	PTC2_HUMAN	Q9y6c5 homo sapien
569	39	45.3	669	2	Q8KKW5_RHIET	Q8kkw5 rhizobium e	642	39	45.3	1203	2	Q53Z57_HUMAN	Q53z57 homo sapien
570	39	45.3	672	2	Q66F41_YERPS	Q66f41 yersinia ps	643	39	45.3	1205	2	Q5SQRS_HUMAN	Q5sq85 homo sapien
571	39	45.3	672	2	Q8ZBE2_YERPE	Q8zbe2 yersinia pe	644	39	45.3	1209	1	KCMW1_MOUSE	Q08460 m calcium-a
572	39	45.3	686	2	Q3G6I3_DDELT	Q3g6i3 pelobacter	645	39	45.3	1209	2	Q5SQR2_HUMAN	Q5sq82 homo sapien
573	39	45.3	696	2	Q784L8_NEUCR	Q784l8 neuropsora	646	39	45.3	1210	1	KCMW1_RAT	Q62976 r calcium-a
574	39	45.3	699	2	Q43TV5_SOLUS	Q43tv5 solibacter	647	39	45.3	1210	2	Q5SQR4_HUMAN	Q5sq84 homo sapien
575	39	45.3	745	2	Q8K1O3_MOUSE	Q8k1o3 mus musculu	648	39	45.3	1222	2	Q7MFT6_VIBVU	Q7mft6 vibrio vuln
576	39	45.3	752	2	Q3I158_PSEHT	Q3i158 pseudoalter	649	39	45.3	1222	2	Q8D4B1_VIBVU	Q8d4b1 vibrio vuln
577	39	45.3	753	2	Q3EHF2_ACTSC	Q3ehf2 actinobacil	650	39	45.3	1225	2	Q5SQS0_HUMAN	Q5sq80 homo sapien
578	39	45.3	756	2	Q2SPI3_GGAMM	Q2spi3 habella che	651	39	45.3	1231	2	Q73730_TRASC	Q73730 trachemys s
579	39	45.3	762	2	Q30368_RALEU	Q30368 ralstonia e	652	39	45.3	1236	1	KCMW1_HUMAN	Q12791 h calcium-a
580	39	45.3	762	2	Q7WX97_RALEU	Q7wx97 ralstonia e	653	39	45.3	1249	2	Q5RFV1_BRARE	Q5rfv1 brachydanio
581	39	45.3	773	2	Q3UQ23_MOUSE	Q3uq23 mus musculu	654	39	45.3	1251	2	Q4S3G4_TETNG	Q4s3g4 tetraodon n
582	39	45.3	807	2	Q6CRP5_YARLI	Q6cfr5 yarrowia li	655	39	45.3	1289	2	Q17174_BOOMI	Q17174 boophilus m
583	39	45.3	809	1	IL4RA_HORSE	Q8wg24 equus cabal	656	39	45.3	1320	2	Q7XXG1_ORYSA	Q7xxg1 oryza sativ
584	39	45.3	826	2	Q2ZMU9_SHEPU	Q2zmu9 shewanella	657	39	45.3	1337	2	Q53KQ8_ORYSA	Q53kq8 oryza sativ
585	39	45.3	826	2	Q366Z1_GAMM	Q366z1 shewanella	658	39	45.3	1428	2	Q7Q867_ANOAG	Q7q867 anopheles g
586	39	45.3	833	2	Q9U2S8_CAEEL	Q9u2s8 caenorhabdi	659	39	45.3	1511	2	Q56A10_MOUSE	Q56a10 mus musculu
587	39	45.3	837	2	Q9HTE2_PSEAE	Q9hte2 pseudomonas	660	39	45.3	1534	2	Q9ULD9_HUMAN	Q9uld9 homo sapien
588	39	45.3	880	2	Q6PLP7_CHLRE	Q6plp7 chlamydomon	661	39	45.3	1756	2	Q6LTS2_CAEBR	Q6lts2 caenorhabdi
589	39	45.3	885	2	Q5JQ19_HUMAN	Q5jq19 homo sapien	662	39	45.3	1801	2	Q2M085_DROPS	Q2m085 drosophila
590	39	45.3	889	2	Q83T68_SALTI	Q83t68 salmonella	663	39	45.3	1919	2	Q29518_RABIT	Q29518 oryctolagus
591	39	45.3	889	2	Q9X6X6_SALMO	Q9x6x6 salmonella	664	39	45.3	1920	2	Q29519_RABIT	Q29519 oryctolagus
592	39	45.3	889	2	Q57MB4_SALCH	Q57mb4 salmonella	665	39	45.3	1926	1	LPH_RABIT	P09848 oryctolagus
593	39	45.3	889	2	Q5PCP1_SALPA	Q5pcp1 salmonella	666	39	45.3	1927	1	Q4ZG58_HUMAN	Q4zg58 homo sapien
594	39	45.3	889	2	Q8Z563_SALTI	Q8z563 salmonella	667	39	45.3	1927	2	Q4HZN3_GIBZE	Q4hzn3 gibberella
595	39	45.3	889	2	Q8ZNH2_SALTY	Q8znh2 salmonella	668	39	45.3	3213	2	Q55934_9HIV1	Q55934 human immun
596	39	45.3	890	1	YOJN_ECOLI	P39838 escherichia	669	38.5	44.8	192	2	Q3QVU5_SILICBACTE	Q3qvus silicibacte
597	39	45.3	890	2	Q31239_SHIBS	Q31239 shigella bo	670	38.5	44.8	295	2	Q3QVU5_9RHOB	Q3qvus bos taurus
598	39	45.3	890	2	Q32119_SHIDS	Q32119 shigella dy	671	38.5	44.8	504	2	Q2TBU6_BOVIN	Q2tbu6 bos taurus
599	39	45.3	890	2	Q3Y220_SHISS	Q3y220 shigella so	672	38.5	44.8	559	2	Q4EFU4_LISMO	Q4efu4 listeria mo
600	39	45.3	890	2	Q7UC77_SHIFL	Q7uc77 shigella fl	673	38.5	44.8	563	2	Q4ESB5_LISMO	Q4eeb5 listeria mo
601	39	45.3	890	2	Q83QW6_SHIFL	Q83qw6 shigella fl	674	38.5	44.8	563	2	Q71ZV4_LISMF	Q71zv4 listeria mo
602	39	45.3	890	2	Q8FPQ0_ECOL6	Q8fpq0 escherichia	675	38.5	44.8	563	2	Q2B77_LISIN	Q2b77 listeria in
603	39	45.3	890	2	Q8XE40_ECO57	Q8xe40 escherichia	676	38.5	44.8	563	2	Q8Y7B8_LISMO	Q8y7b8 listeria mo
604	39	45.3	910	2	Q3XLH5_PROT	Q3xlh5 magnetococc	677	38.5	44.8	609	1	LKHA4_RAT	P30349 rattus norv
605	39	45.3	913	2	Q4N589_THERA	Q4n589 theileria p	678	38.5	44.8	610	1	LKHA4_CAVPO	P19602 cavia porce
606	39	45.3	915	2	Q75G97_ORYSA	Q75g97 oryza sativ	679	38.5	44.8	610	1	LKHA4_CHILA	Q689c8 chinchilla
607	39	45.3	924	2	Q2QRT7_ORYSA	Q2qrt7 oryza sativ	680	38.5	44.8	610	1	LKHA4_HUMAN	P09960 homo sapien
608	39	45.3	925	2	Q5ARA5_EMENI	Q5ara5 aspergillus	681	38.5	44.8	610	1	LKHA4_MOUSE	F24527 mus musculu
609	39	45.3	952	2	Q6Q0N0_RAT	Q6q0n0 rattus norv	682	38.5	44.8	611	2	Q6IAT6_HUMAN	Q6iat6 homo sapien
610	39	45.3	959	2	Q7TS67_MOUSE	Q7ts67 mus musculu	683	38.5	44.8	611	2	Q5REQ3_PONPY	Q5req3 pongo pygma
611	39	45.3	976	2	Q6FKX5_CANGA	Q6fkx5 candida gla	684	38.5	44.8	611	2	Q3SZH7_BOVIN	Q3szh7 bos taurus
612	39	45.3	979	1	CSTNI_MOUSE	Q9ep12 mus musculu	685	38.5	44.8	611	2	Q3UY71_MOUSE	Q3uy71 mus musculu
613	39	45.3	979	2	Q3HGG9_TRIER	Q3hgg9 trichodeemi	686	38.5	44.8	611	2	Q499P2_RAT	Q499p2 rattus norv
614	39	45.3	994	2	Q5SLY5_THET8	Q5sly5 thermus the	687	38	44.2	36	2	Q4XFD7_PLACH	Q4xfv7 plasmodium
615	39	45.3	994	2	Q72GN0_THET2	Q72gm0 thermus the	688	38	44.2	48	2	Q855D7_9CAUD	Q855d7 mycobacteri

689	38	44.2	49	2	Q855N4_9CAUD	Q855N4 mycobacteri	762	38	44.2	289	1	ATLSL_ARATH	Q9fir0 arabidopsis
690	38	44.2	68	2	Q6K2B5_ORYSA	Q6K2B5 oryza sativ	763	38	44.2	298	2	Q7XAT5_VICFA	Q7xat5 vicia faba
691	38	44.2	68	2	Q3EVG7_BACTI	Q3evg7 bacillus th	764	38	44.2	301	2	Q4K982_PSEFA	Q4k982 pseudomonas
692	38	44.2	88	2	Q56F07_9CAUD	Q56f07 aeromonas p	765	38	44.2	301	2	Q6MLG1_BDEBA	Q6mlg1 bdellovibri
693	38	44.2	88	2	Q6U9U8_9CAUD	Q6u9u8 aeromonas p	766	38	44.2	308	2	Q49X95_STAS1	Q49x95 staphylococ
694	38	44.2	89	2	Q8GC20_LEUME	Q8gc20 leuconostoc	767	38	44.2	309	2	Q54G71_DICDI	Q54g71 dictyosteli
695	38	44.2	105	2	Q6PSX5_HUMAN	Q6psx5 homo sapien	768	38	44.2	310	2	Q8XT83_RALSO	Q8xt83 raistonia s
696	38	44.2	105	2	Q5RE30_PONPY	Q5re30 pongo pygma	769	38	44.2	314	2	Q54DH5_DICDI	Q54dh5 dictyosteli
697	38	44.2	107	2	Q5Z5F0_ORYSA	Q5z5f0 oryza sativ	770	38	44.2	317	2	Q93YE3_TOBAC	Q93ye3 nicotiana t
698	38	44.2	110	2	Q31Q08_SYNP7	Q31q08 synechococc	771	38	44.2	319	2	Q2JC25_9ACTO	Q2jc25 frankia sp.
699	38	44.2	114	2	Q8Q6E9_9H1V1	Q8q6e9 human immu	772	38	44.2	324	2	Q7NPK5_GLOVI	Q7npk5 gloeobacter
700	38	44.2	115	2	Q7ZC25_9H1V1	Q7zc25 human immu	773	38	44.2	325	2	Q68334_VIBCH	Q68334 vibrio chol
701	38	44.2	121	2	Q8CC75_MOUSE	Q8cc75 mus musculu	774	38	44.2	325	2	Q8D414_VIBVU	Q8d414 vibrio vuln
702	38	44.2	136	2	Q7UQE9_RHOBA	Q7uqe9 rhodopirell	775	38	44.2	326	2	Q3QER6_9CMM	Q3qer6 shewanella
703	38	44.2	136	2	Q9X8X1_STRCO	Q9x8x1 streptomyce	776	38	44.2	327	2	Q9KTS1_VIBCH	Q9kts1 vibrio chol
704	38	44.2	137	2	Q4CJ34_CLOTH	Q4cj34 clostridium	777	38	44.2	334	2	Q5ND67_ORYLA	Q5nd67 oryzias lat
705	38	44.2	144	2	Q2ZD10_9GANN	Q2zdi0 shewanella	778	38	44.2	335	2	Q2VWQ7_ASPFU	Q2vwq7 aspergillus
706	38	44.2	144	2	Q8RGG9_FUSNN	Q8rgg9 fusobacteri	779	38	44.2	335	2	Q4WIX5_ASPFU	Q4wix5 aspergillus
707	38	44.2	150	2	Q5LLM9_SILFO	Q5llm9 elicibacte	780	38	44.2	340	2	Q5OJ70_BRARE	Q5oj70 brachydanio
708	38	44.2	162	2	Q7F999_ORYSA	Q7f999 oryza sativ	781	38	44.2	340	2	Q561V1_BRARE	Q561v1 brachydanio
709	38	44.2	162	2	Q7XR88_ORYSA	Q7xr88 oryza sativ	782	38	44.2	341	2	Q6EV32_BRARE	Q6ev32 brachydanio
710	38	44.2	164	2	Q7P4Z6_FUSNV	Q7p4z6 fusobacteri	783	38	44.2	342	2	Q33HD4_METHU	Q33hd4 methanospir
711	38	44.2	164	2	Q3C4C0_9CLOT	Q3c4c0 alkaliphilu	784	38	44.2	342	2	Q6NN16_DROME	Q6nn16 drosophila
712	38	44.2	169	2	Q2TIF9_ASPOR	Q2tif9 aspergillus	785	38	44.2	342	2	Q8INF2_DROME	Q8inf2 drosophila
713	38	44.2	170	2	Q7QG36_ANOGA	Q7qg36 anopheles g	786	38	44.2	343	2	Q2UQT9_ASPOR	Q2uqt9 aspergillus
714	38	44.2	172	2	Q4NSP5_THEPA	Q4nsp5 theileria p	787	38	44.2	344	2	Q32YJ9_9EUKA	Q32yj9 ammonia sp.
715	38	44.2	181	2	Q4UDU9_THEAN	Q4udu9 theileria a	788	38	44.2	345	2	Q51932_PEPWA	Q51932 peptostrept
716	38	44.2	182	2	Q5CIZ1_CRYHO	Q5ciz1 cryptospori	789	38	44.2	350	2	Q5OJ63_FUGRU	Q5oj63 fugu rubrip
717	38	44.2	182	2	Q8DSR4_STRWJ	Q8dsr4 streptococc	790	38	44.2	350	2	Q5TIN7_TETNG	Q5tin7 tetraodon n
718	38	44.2	183	2	Q7QVN8_GIALA	Q7qvn8 giardia lam	791	38	44.2	351	2	Q702S1_FUGRU	Q702s1 fugu rubrip
719	38	44.2	185	2	Q9VE56_DROME	Q9ve56 drosophila	792	38	44.2	352	1	DHSO_BACSU	Q60004 bacillus su
720	38	44.2	186	1	ARL6_HUMAN	Q9h0f7 homo sapien	793	38	44.2	356	2	Q82B48_STRAW	Q82b48 streptomyce
721	38	44.2	186	1	ARL6_PONPY	Q5r4q5 pongo pygma	794	38	44.2	357	2	Q31E22_THICR	Q31e22 thiomicrosp
722	38	44.2	186	2	Q5M9P8_BRARE	Q5m9p8 brachydanio	795	38	44.2	358	2	Q4KMS4_HUMAN	Q4kms4 homo sapien
723	38	44.2	187	2	Q4CCK1_CLOTH	Q4cck1 clostridium	796	38	44.2	361	2	Q5US34_BRARE	Q5us34 brachydanio
724	38	44.2	188	2	Q49WF9_STAS1	Q49wf9 staphylococ	797	38	44.2	363	2	Q3FYD7_9DELT	Q3fyd7 pelobacter
725	38	44.2	188	2	Q4SI06_TETNG	Q4si06 tetraodon n	798	38	44.2	365	2	Q9KXK4_STRCO	Q9kxk4 streptomyce
726	38	44.2	193	2	Q68EX8_XENLA	Q68ex8 xenopus lae	799	38	44.2	370	2	Q58M48_9CAUD	Q58m48 cyanophag
727	38	44.2	194	2	Q6CC85_YARLI	Q6cc85 varrowia li	800	38	44.2	370	2	Q4CJL3_CLOTH	Q4cj13 clostridium
728	38	44.2	201	2	Q4A2V2_9PHYC	Q4a2v2 emiliania h	801	38	44.2	374	2	Q6EV31_BRARE	Q6ev31 brachydanio
729	38	44.2	209	2	Q3FEI3_9BURK	Q3fe13 burkholderi	802	38	44.2	374	2	Q5ND65_ORYLA	Q5nd65 oryzias lat
730	38	44.2	212	2	Q3RSK6_RALME	Q3rsk6 ralstonia m	803	38	44.2	375	1	ACT_GIALA	P51775 giardia lam
731	38	44.2	212	2	Q88TQ8_LACPL	Q88tq8 lactobacill	804	38	44.2	375	2	Q2U7A3_ASPOR	Q2u7a3 aspergillus
732	38	44.2	214	2	Q3FRJ6_9BURK	Q3frj6 lactoferax	805	38	44.2	375	2	Q7QVY2_GIALA	Q7qv2 giardia lam
733	38	44.2	217	2	Q9KMV8_VIBCH	Q9kmv8 vibrio chol	806	38	44.2	380	2	Q8TOK7_DROME	Q8tok7 drosophila
734	38	44.2	220	2	Q8SU59_ENCCU	Q8su59 encephalito	807	38	44.2	381	2	Q74D11_GEOSL	Q74d11 geobacter s
735	38	44.2	220	2	Q3ERY5_BACTI	Q3ery5 bacillus th	808	38	44.2	384	2	Q93YE7_TOBAC	Q93ye7 nicotiana t
736	38	44.2	224	2	Q53IY4_RHOS2	Q53iy4 rhodomonas	809	38	44.2	384	2	Q93YF0_TOBAC	Q93yf0 nicotiana t
737	38	44.2	224	2	Q9LKN8_GUITH	Q9lkn8 guillardia	810	38	44.2	385	1	YG1W_YEAST	P53230 saccharomyc
738	38	44.2	225	2	Q9QDB1_9GEMI	Q9qdb1 cowpea gold	811	38	44.2	385	2	Q45U38_YEAST	Q45u38 saccharomyc
739	38	44.2	226	2	Q3VIB6_9CHLB	Q3vib6 pelodictyon	812	38	44.2	391	2	Q2UQR1_ASPOR	Q2uqr1 aspergillus
740	38	44.2	227	2	Q3YJ63_9LILI	Q3yj63 euterge ole	813	38	44.2	396	2	Q82QT0_STRAW	Q82qt0 streptomyce
741	38	44.2	233	2	Q50J71_BRARE	Q50j71 brachydanio	814	38	44.2	398	2	Q2ZL10_SHEPU	Q2z10 shewanella
742	38	44.2	246	2	Q499B8_BRARE	Q499b8 brachydanio	815	38	44.2	402	1	2322A_HUMAN	Q2110 hom sapien
743	38	44.2	248	2	Q68SX1_9CIRC	Q68sx1 beak and fe	816	38	44.2	402	1	2322A_MACFA	Q4r7x8 macaca fasc
744	38	44.2	255	2	Q4V208_BACCZ	Q4v208 bacillus ce	817	38	44.2	402	2	Q4TIH3_TETNG	Q4t1h3 tetraodon n
745	38	44.2	256	2	Q3FH96_9BURK	Q3fh96 burkholderi	818	38	44.2	404	2	Q871K2_NEUCR	Q871k2 neurospora
746	38	44.2	256	2	Q3J9721_BURSD	Q3j9721 burkholderi	819	38	44.2	410	1	2322A_MOUSE	Q8bz89 mus musculu
747	38	44.2	256	2	Q32234_BACSU	Q32234 bacillus su	820	38	44.2	410	2	Q3TDD5_MOUSE	Q3tdd5 mus musculu
748	38	44.2	259	2	Q34439_BACSU	Q34439 bacillus su	821	38	44.2	419	1	RBM4_BRARE	Q6iq97 brachydanio
749	38	44.2	267	1	YBX1_BACSU	P54427 bacillus su	822	38	44.2	420	2	Q8EL05_OCEIH	Q8el05 oceanobacil
750	38	44.2	269	2	Q53FD4_HUMAN	Q53fd4 homo sapien	823	38	44.2	423	2	Q49A28_HUMAN	Q49a28 homo sapien
751	38	44.2	269	2	Q2KSD7_ADE04	Q2ksd7 human adeno	824	38	44.2	425	2	Q93YD0_TOBAC	Q93yd0 nicotiana t
752	38	44.2	274	2	Q6H1B3_9ADEN	Q6h1b3 human adeno	825	38	44.2	425	2	Q93YD6_TOBAC	Q93yd6 nicotiana t
753	38	44.2	274	2	Q8BEL2_ADE04	Q8bel2 human adeno	826	38	44.2	425	2	Q9XG72_TOBAC	Q9xg72 nicotiana t
754	38	44.2	276	2	Q2STU2_BURTH	Q2stu2 burkholderi	827	38	44.2	425	2	Q93YE0_TOBAC	Q93ye0 nicotiana t
755	38	44.2	276	2	Q3JV81_BURPL	Q3jv81 burkholderi	828	38	44.2	426	2	Q93YE0_TOBAC	Q93ye0 nicotiana t
756	38	44.2	276	2	Q6ZG24_BURMA	Q6zg24 burkholderi	829	38	44.2	438	2	Q2J527_9ACTO	Q2j527 frankia sp.
757	38	44.2	276	2	Q63PT4_BURES	Q63pt4 burkholderi	830	38	44.2	439	2	Q67K70_SYMTH	Q67k70 symbiobacte
758	38	44.2	283	2	Q583W4_9TRYP	Q583w4 trypanosoma	831	38	44.2	442	2	Q4S1E0_TETNG	Q4s1e0 tetraodon n
759	38	44.2	285	2	Q4KN40_HUMAN	Q4kn40 homo sapien	832	38	44.2	450	2	Q9K9L6_BACHD	Q9k9l6 bacillus ha
760	38	44.2	285	2	Q9H9F7_HUMAN	Q9h9f7 homo sapien	833	38	44.2	452	2	Q7NMK8_GLOVI	Q7nmk8 gloeobacter
761	38	44.2	288	2	Q9ZXA4_BPPHC	Q9zxa4 bacterioph	834	38	44.2	453	2	Q37ZF4_SPHAR	Q37zf4 novosphingo



835	38	44.2	456	2	Q3EVW8_BACTI	Q3evw8 bacillus th	908	38	44.2	898	2	Q7MUA9_PORGI	Q7mu9 porphyromon
836	38	44.2	469	2	Q7PYE2_ANOGA	Q7pye2 anopheles g	909	38	44.2	901	2	Q9V172_DROME	Q9v172 drosophila
837	38	44.2	481	2	Q6WJ8_SCHMA	Q6wj8 schistosoma	910	38	44.2	904	2	Q2SDG8_ORYSA	Q2sdg8 oryza sativ
838	38	44.2	481	2	Q7NA15_MYCGA	Q7na15 mycoplasma	911	38	44.2	906	2	Q4Z052_PLABE	Q4z052 plasmodium
839	38	44.2	483	2	Q4PHK1_USTMA	Q4phk1 ustilago ma	912	38	44.2	909	2	Q6FLK8_CANGA	Q6flk8 candida gla
840	38	44.2	487	2	Q76399_CABEL	Q76399 caenorhabdi	913	38	44.2	933	2	Q4H9L2_9DEIO	Q4h9l2 deinococcus
841	38	44.2	488	2	Q8A510_BACTN	Q8a510 bacteroides	914	38	44.2	964	1	CTA2_BACCI	P70873 bacillus ci
842	38	44.2	491	2	Q5LM30_STRT1	Q5lm30 streptococc	915	38	44.2	980	2	Q982K5_RHILO	Q982k5 rhizobium l
843	38	44.2	491	2	Q5M393_STRT2	Q5m393 streptococc	916	38	44.2	997	2	Q4UDD9_THEAN	Q4udd9 theileria a
844	38	44.2	499	2	Q2NQB1_SODGL	Q2nqb1 sodalis glo	917	38	44.2	1007	2	Q376P3_RHOPA	Q376p3 rhodopseudo
845	38	44.2	503	2	Q2U7B5_ASPOR	Q2u7b5 aspergillus	918	38	44.2	1011	2	Q2UCZ1_ASPOR	Q2ucz1 aspergillus
846	38	44.2	503	2	Q44T25_CHLLI	Q44t25 chlorobium	919	38	44.2	1016	2	Q44AJ5_SOLUS	Q44aj5 solibacter
847	38	44.2	508	2	Q8W37_LACPL	Q8w37 lactobacill	920	38	44.2	1047	2	Q8IPB9_DROME	Q8ipb9 drosophila
848	38	44.2	517	2	Q2N587_SPHN	Q2n5e7 erythrobact	921	38	44.2	1051	2	Q9VKX1_DROME	Q9vkk1 drosophila
849	38	44.2	519	2	Q49AM3_HUMAN	Q49am3 homo sapien	922	38	44.2	1055	2	Q5OX17_ENTHI	Q5oxi7 entamoeba h
850	38	44.2	520	2	Q7VIB9_HELHP	Q7vib9 helicobacte	923	38	44.2	1103	2	Q5QXB7_HUMAN	Q5qxb7 entamoeba h
851	38	44.2	521	2	Q4WM02_ASPFU	Q4wm02 aspergillus	924	38	44.2	1139	1	MA2A2_HUMAN	Q3gbd2 synchromo
852	38	44.2	525	2	Q5LN01_STLPO	Q5ln01 silicibacte	925	38	44.2	1182	2	Q7UJC2_RHOBA	P49641 homo sapien
853	38	44.2	528	2	Q3FW33_BURK	Q3fw33 rhodofera	926	38	44.2	1187	2	Q6NSM8_BRARE	Q7ujc2 rhodopirell
854	38	44.2	529	2	Q6QN29_NPVSP	Q6qn29 spidoptera	927	38	44.2	1189	2	Q6PHV1_BRARE	Q6ns8 brachydania
855	38	44.2	531	2	Q74131_YARLI	Q74131 yarrowia li	928	38	44.2	1233	2	Q75H40_ORYSA	Q6phv1 brachydania
856	38	44.2	534	2	Q97XY4_SULSO	Q97xy4 sulfolobus	929	38	44.2	1238	2	Q5W6J6_ORYSA	Q75h40 oryza sativ
857	38	44.2	538	1	CP392_DROME	P82713 drosophila	930	38	44.2	1305	2	Q4SPN5_TETNG	Q5w6j6 oryza sativ
858	38	44.2	541	2	Q7PR27_ANOGA	Q7pr27 anopheles g	931	38	44.2	1326	2	Q8AW94_FUGRU	Q4spn5 tetraodon n
859	38	44.2	554	2	Q4T606_TETNG	Q4t606 tetraodon n	932	38	44.2	1346	2	Q3HA57_TRIER	Q8aw94 fugu rubrin
860	38	44.2	566	2	P77943_SULAC	P77943 sulfolobus	933	38	44.2	1437	2	Q5S171_CRYNE	Q3ha57 trichodesmi
861	38	44.2	566	2	Q4JA37_SULAC	Q4ja37 sulfolobus	934	38	44.2	1437	2	Q5K7Q6_CRYNE	Q5s171 cryptococcu
862	38	44.2	575	2	Q8RID9_TETNG	Q8rid9 tetraodon n	935	38	44.2	1470	2	Q3J8W4_NITOC	Q5k7q6 cryptococcu
863	38	44.2	591	2	Q8AAV1_BACTN	Q8aav1 bacteroides	936	38	44.2	1573	2	Q3KDT3_PSEPF	Q3j8w4 nitrosococc
864	38	44.2	604	2	Q53PG0_ORYSA	Q53pg0 oryza sativ	937	38	44.2	1604	2	Q7XQ14_ORYSA	Q3kdt3 pseudomonas
865	38	44.2	609	2	Q3MY84_9DELT	Q3my84 syntrophoba	938	38	44.2	1845	2	Q8OUA8_MOUSE	Q7xq14 oryza sativ
866	38	44.2	612	2	Q8MU94_MUSDO	Q8mu94 musca domes	939	38	44.2	1845	2	Q5QNG3_MOUSE	Q8oua8 mus musculu
867	38	44.2	629	2	Q7QFG0_ANOGA	Q7qfg0 anopheles g	940	38	44.2	1855	2	Q8QZFO_RAT	Q5qng3 mus musculu
868	38	44.2	631	2	Q8EJ70_SHEON	Q8ej70 shewanella	941	38	44.2	1860	2	Q8IZC6_HUMAN	Q8qzf0 rattus norv
869	38	44.2	633	2	Q75VY0_9DIPT	Q75vy0 culicx trita	942	38	44.2	2095	2	Q4RYE5_TETNG	Q8izc6 homo sapien
870	38	44.2	634	2	Q7UP15_RHOBA	Q7up15 rhodopirell	943	38	44.2	2174	2	Q2GQRO_DROME	Q4rye5 tetraodon n
871	38	44.2	638	2	Q3Q0K3_GGAMM	Q3q0k3 shewanella	944	38	44.2	2348	2	Q36WE4_RHOPA	Q2gqro drosophila
872	38	44.2	641	2	Q2X2K3_GGAMM	Q2x2k3 shewanella	945	38	44.2	2772	2	Q3AVAV_DROME	Q36we4 rhodopseudo
873	38	44.2	644	2	Q2ZB70_GGAMM	Q2zb70 shewanella	946	38	44.2	2776	2	Q869AO_DROME	Q3avav drosophila
874	38	44.2	644	2	Q363P1_GGAMM	Q363p1 shewanella	947	38	44.2	2894	2	Q7KRX2_DROME	Q869ao drosophila
875	38	44.2	645	2	Q7RTL9_ANOGA	Q7rtl9 anopheles g	948	38	44.2	2898	2	Q868Z9_DROME	Q7krx2 drosophila
876	38	44.2	647	2	Q3ARE9_CHLCH	Q3are9 chlorobium	949	37.5	43.6	147	1	RS12_METWP	Q6lxi4 methanococc
877	38	44.2	648	2	Q36HC5_GGAMM	Q36hcs shewanella	950	37.5	43.6	162	2	Q9S394_LACHI	Q6lxi4 lactobacill
878	38	44.2	649	1	ACES_DROME	P07140 drosophila	951	37.5	43.6	192	2	Q5S994_9HIV1	Q9s394 human immun
879	38	44.2	650	2	Q5TS70_ANOGA	Q5ts70 anopheles g	952	37.5	43.6	192	2	Q4JGK4_HUMAN	Q5s994 human immun
880	38	44.2	658	2	Q6CD14_YARLI	Q6cd14 yarrowia li	953	37.5	43.6	192	2	Q5EEI6_9HIV1	Q4jgk4 human immun
881	38	44.2	664	1	ACES_ANOST	P56161 anopheles s	954	37.5	43.6	192	2	Q8UTS6_9HIV1	Q5eei6 human immun
882	38	44.2	673	2	Q5KIT3_BACDO	Q5ktt3 bactrocera	955	37.5	43.6	192	2	Q2VAK8_9HIV1	Q8uts6 human immun
883	38	44.2	673	2	Q5QTL8_BACDO	Q5qtl8 bactrocera	956	37.5	43.6	199	1	SODM2_HALVO	Q2vak8 human immun
884	38	44.2	673	2	Q8MVZ4_BACOL	Q8mvz4 bactrocera	957	37.5	43.6	200	1	SODM1_HALVO	Q03301 halobacteri
885	38	44.2	678	2	Q5BF27_EMENI	Q5bf27 aspergillus	958	37.5	43.6	269	2	Q5CZ70_HUMAN	Q03300 halobacteri
886	38	44.2	681	2	Q8ZV43_PYRAE	Q8zvx3 pyrobaculum	959	37.5	43.6	293	2	Q7NTE7_CHRVO	Q5cz70 homo sapien
887	38	44.2	691	2	Q8MXC5_MUSDO	Q8mx5 musca domes	960	37.5	43.6	307	2	Q894D9_CLOTE	Q7nte7 chromobacte
888	38	44.2	691	2	Q2YHQ7_MUSDO	Q2yhq7 musca domes	961	37.5	43.6	310	2	Q6RXB1_HCMV	Q894d9 clostridium
889	38	44.2	692	2	Q7YMW9_MUSDO	Q7ymw9 musca domes	962	37.5	43.6	310	2	Q6SVZ2_HCMV	Q6rxb1 human cytom
890	38	44.2	692	2	Q8MXC4_MUSDO	Q8mx4 musca domes	963	37.5	43.6	329	2	Q337Y9_ORYSA	Q6svz2 human cytom
891	38	44.2	692	2	Q8MXC6_MUSDO	Q8mx6 musca domes	964	37.5	43.6	362	2	Q7XEE0_ORYSA	Q337y9 oryza sativ
892	38	44.2	692	2	Q8MXC7_MUSDO	Q8mx7 musca domes	965	37.5	43.6	385	2	Q645R2_TARGR	Q7xee0 oryza sativ
893	38	44.2	692	2	Q8MXC8_MUSDO	Q8mx8 musca domes	966	37.5	43.6	424	2	Q6SE84_9CAUD	Q645r2 taricha gra
894	38	44.2	692	2	Q8MXC9_MUSDO	Q8mx9 musca domes	967	37.5	43.6	424	2	Q6SPU5_LACJO	Q6se84 lactobacill
895	38	44.2	692	2	Q9SP20_MUSDO	Q9sp20 musca domes	968	37.5	43.6	440	2	Q6X2M1_9BIVA	Q6spu5 lactobacill
896	38	44.2	692	2	Q9SWV7_MUSDO	Q9swv7 musca domes	969	37.5	43.6	466	2	Q6ZQH9_MOUSE	Q6x2m1 chlamys far
897	38	44.2	704	2	Q3BDV8_HAEIR	Q3bdv8 haematobia	970	37.5	43.6	478	2	Q8C2R4_MOUSE	Q6zqh9 mus musculu
898	38	44.2	708	2	P91954_LUCCU	P91954 lucilia cup	971	37.5	43.6	546	2	Q3TO62_MOUSE	Q8c2r4 mus musculu
899	38	44.2	714	2	Q5T1U7_HUMAN	Q5t1u7 homo sapien	972	37.5	43.6	546	2	Q66HD9_RAT	Q3to62 mus musculu
900	38	44.2	730	2	Q4WNP3_ASPFU	Q4wnp3 aspergillus	973	37.5	43.6	548	1	IDD_MOUSE	Q66hd9 rattus norv
901	38	44.2	756	2	Q39PX6_GEOGM	Q39px6 geobacter m	974	37.5	43.6	549	2	Q6P5A9_MOUSE	P98154 mus musculu
902	38	44.2	773	2	Q93FB8_9RHOM	Q93fb8 azoarcus ev	975	37.5	43.6	550	1	IDD_HUMAN	Q6p5a9 m digeorge
903	38	44.2	774	2	Q9IAV5_BRARE	Q9iav5 brachydania	976	37.5	43.6	550	2	Q8IWC8_HUMAN	P98153 homo sapien
904	38	44.2	796	2	Q604T9_NETCA	Q604t9 methylococc	977	37.5	43.6	550	2	Q5R6P9_PONPY	Q8iwc8 homo sapien
905	38	44.2	798	1	HASP_HUMAN	Q8ctf76 homo sapien	978	37.5	43.6	554	2	Q21P51_9DELT	Q5r6p9 pongo pygma
906	38	44.2	812	2	Q9IAV4_BRARE	Q9iav4 brachydania	979	37.5	43.6	565	2	Q3F388_RHOMR	Q21p51 anaeromyxob
907	38	44.2	887	2	Q3WPP2_9RHIZ	Q3wpp2 mesorhizobi	980	37.5	43.6	580	2	Q8CB23_MOUSE	Q9f388 rhodothermu



981 37.5 43.6 588 2 Q3Z7M9\_DEHE1 Q3Z7m9 dehalococco  
 982 37.5 43.6 1081 2 Q6F8U9\_ACIDR Q6f8u9 acinetobact  
 983 37.5 43.6 1241 2 Q80S25\_NDIAD Q80s25 ndum virus  
 984 37.5 43.6 1248 2 Q80S39\_9VIRU Q80s39 bebaru viru  
 985 37.5 43.6 1266 2 Q9LAA1\_STASI Q9laa1 staphylococ  
 986 37 43.0 33 2 Q4RIH4\_PIG Q4rli4 sus scrofa  
 987 37 43.0 47 2 Q70U88\_9AGAR Q70u88 hebeloma ra  
 988 37 43.0 58 2 O18895\_CANFA O18895 canis famil  
 989 37 43.0 59 2 Q37CK6\_RHOPA Q37ck6 rhodopseudo  
 990 37 43.0 59 2 Q2IRN0\_RHOPA Q2irn0 rhodopseudo  
 991 37 43.0 76 2 Q5YVQ0\_NOCPA Q5yvw0 nocardia fa  
 992 37 43.0 86 1 VFW\_CMVVS P19510 ovine lent  
 993 37 43.0 97 2 Q4Z9Q0\_9CAUD Q4z9q0 bacterioph  
 994 37 43.0 103 2 Q4XJH8\_PLACH Q4xjh8 plasmodium  
 995 37 43.0 107 2 Q4RIH5\_PIG Q4rli5 sus scrofa  
 996 37 43.0 108 2 Q5ALN4\_CANAL Q5aln4 candida alb  
 997 37 43.0 109 2 Q6C043\_YARLI Q6c043 yarrowia li  
 998 37 43.0 110 2 Q38E29\_9TRYP Q38e29 trypanosoma  
 999 37 43.0 113 2 Q77162\_ENTHI Q77162 entamoeba h  
 1000 37 43.0 113 2 Q77163\_ENTIV Q77163 entamoeba i

## ALIGNMENTS

RESULT 1  
 HPSE HUMAN  
 ID HPSE HUMAN STANDARD; PRT; 543 AA.  
 AC Q9Y5E1; Q5G6S; Q9UL39;  
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
 DT 01-NOV-1999, sequence version 1.  
 DT 07-FEB-2006, entry version 27.  
 DE Heparanase precursor (EC 3.2.-.-) (Heparanase-1) (Hpal) (Endo-  
 DE glucuronidase) [Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa  
 DE subunit].  
 GN Name=HPSE; Synonyms=HEP, HPA, HPA1, HPRL, HPSE1, HSE1;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.  
 RC TISSUE=Placenta;  
 RX MEDLINE=99335379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;  
 RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C.,  
 RA Seddon A.P., Giorgio N.A., Bohlen P.;  
 RT "Cloning and functional expression of a human heparanase gene.";  
 RL Biochem. Biophys. Res. Commun. 261:183-187(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN  
 RP SEQUENCE OF 158-168; 326-337 AND 447-491.  
 RC TISSUE=Embryonic fibroblast;  
 RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;  
 RA Toyoshima M., Nakajima M.;  
 RT "Human heparanase. Purification, characterization, cloning, and  
 RT expression.";  
 RL J. Biol. Chem. 274:24153-24160(1999).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.  
 RX PubMed=10395325; DOI=10.1038/10518;  
 RA Vlodavsky I., Friedmann Y., Elkin M., Aingorn H., Atzmon R.,  
 RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,  
 RA Spector L., Pecker I.;  
 RT "Mammalian heparanase: gene cloning, expression and function in tumor  
 RT progression and metastasis.";  
 RL Nat. Med. 5:793-802(1999).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE  
 RP OF 158-174; 263-272; 326-337; 433-436; 466-468 AND 478-483.  
 RC TISSUE=Placenta;  
 RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;  
 RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,

Parish C.R.;  
 RT "Cloning of mammalian heparanase, an important enzyme in tumor  
 RT invasion and metastasis.";  
 RL Nat. Med. 5:803-809(1999).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE  
 RP SPECIFICITY.  
 RC TISSUE=Placenta;  
 RX MEDLINE=2029546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;  
 RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;  
 RT "Heparanase expression in invasive trophoblasts and acute vascular  
 RT damage.";  
 RL Glycobiology 10:467-475(2000).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.  
 RX PubMed=11547900; DOI=10.1023/A:1011375624902;  
 RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O.,  
 RA Aviv A., Elkin M., Pecker I., Peretz T., Vlodavsky I.;  
 RT "Molecular properties and involvement of heparanase in cancer  
 RT progression and mammary gland morphogenesis.";  
 RL J. Mammary Gland Biol. Neoplasia 6:311-322(2001).  
 RN [7]  
 RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163,  
 RP SUBUNITS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.  
 RC TISSUE=Placenta;  
 RX PubMed=12713442; DOI=10.1042/BJ20030318;  
 RA McKenzie E., Young K., Hircock M., Bennett J., Bhaman M., Felix R.,  
 RA Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S.,  
 RA Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M.,  
 RA Parekh R., Stuberfield C.;  
 RT "Biochemical characterization of the active heterodimer form of human  
 RT heparanase (Hpal) protein expressed in insect cells.";  
 RL Biochem. J. 373:423-435(2003).  
 RN [8]  
 RP NUCLEOTIDE SEQUENCE [MRNA].  
 RA Pinhal M.A., Semedo P.;  
 RT "Cloned heparanase from MCF-7 cells.";  
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.  
 RN [9]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
 RC TISSUE=Small intestine;  
 RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,  
 RA Tanaka A., Yokoyama S.;  
 RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.  
 RN [10]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
 RC TISSUE=Pancreas;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Colling F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Stalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [11]  
 RP MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.  
 RX PubMed=11123890; DOI=10.1021/bi002080p;  
 RA Hulett M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,  
 RA Greedy J.E., Parish C.R.;

RT "Identification of active-site residues of the pro-metastatic  
 RT endoglycosidase heparanase.";  
 RL Biochemistry 39:15659-15667(2000).  
 RN [12]  
 RP N-GLYCOSYLATION, AND MUTAGENESIS OF ASN-162; ASN-178; ASN-200;  
 RP ASN-217; ASN-238 AND ASN-459.  
 RX PubMed=14573609; DOI=10.1074/jbc.M300541200;  
 RA Samizu S., Ishida K., Wierzb M.K., Osada H.;  
 RT "Secretion of heparanase protein is regulated by glycosylation in  
 RT human tumor cell lines.";  
 RL J. Biol. Chem. 279:2697-2703(2004).  
 RN [13]  
 RP SUBCELLULAR LOCATION.  
 RX PubMed=15292202; DOI=10.1074/jbc.M402131200;  
 RA Gingis-Velitski S., Zetser A., Kaplan V., Ben-Zaken O., Cohen E.,  
 RA Levy-Adam F., Bashenko Y., Flugelman M.Y., Vlodavsky I., Ilan N.;  
 RT "Heparanase uptake is mediated by cell membrane heparan sulfate  
 RT proteoglycans.";  
 RL J. Biol. Chem. 279:44084-44092(2004).  
 RN [14]  
 RP BIOPHYSICOCHEMICAL PROPERTIES, PROCESSING, AND SUBCELLULAR LOCATION.  
 RX PubMed=15948168; DOI=10.1016/j.febslet.2005.03.030;  
 RA Cohen E., Atzmon R., Vlodavsky I., Ilan N.;  
 RT "Heparanase processing by lysosomal/endosomal protein preparation.";  
 RL FEBS Lett. 579:2334-2338(2005).  
 RN [15]  
 RP SUBCELLULAR LOCATION, PROCESSING, AND MUTAGENESIS OF TYR-156.  
 RX PubMed=15659389; DOI=10.1074/jbc.M413370200;  
 RA Abboud-Jarrour G., Rangini-Guetta Z., Aingorn H., Atzmon R.,  
 RA Elgavish S., Peretz T., Vlodavsky I.;  
 RT "Site-directed mutagenesis, proteolytic cleavage, and activation of  
 RT human proheparanase.";  
 RL J. Biol. Chem. 280:13568-13575(2005).  
 RN [16]  
 RP DOMAINS, AND MUTAGENESIS OF LYS-158 AND LYS-161.  
 RX PubMed=15769092; DOI=10.1074/jbc.M414546200;  
 RA Levy-Adam F., Abboud-Jarrour G., Guerrini M., Beccati D.,  
 RA Vlodavsky I., Ilan N.;  
 RT "Identification and characterization of heparin/heparan sulfate  
 RT binding domains of the endoglycosidase heparanase.";  
 RL J. Biol. Chem. 280:20457-20466(2005).  
 RN [17]  
 RP VARIANT SER-260.  
 RX PubMed=15334672;  
 RA Chen X.P., Liu Y.B., Rui J., Peng S.Y., Peng C.H., Zhou Z.Y.,  
 RA Shi L.H., Shen H.W., Xu B.;  
 RT "Heparanase mRNA expression and point mutation in hepatocellular  
 RT carcinoma.";  
 RL World J. Gastroenterol. 10:2795-2799(2004).  
 CC -1- FUNCTION: Endoglycosidase which is a cell surface and  
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
 CC proteoglycans. Also implicated in the extravasation of leukocytes  
 CC and tumor cell lines. Due to its contribution to metastasis and  
 CC angiogenesis, it is considered to be a potential target for anti-  
 CC cancer therapies.  
 CC -1- ENZYME REGULATION: Inhibited by EDTA, laminarin sulfate and, to a  
 CC lower extent, by heparin and sulfamin and activated by calcium and  
 CC magnesium (By similarity).  
 CC -1- BIOPHYSICOCHEMICAL PROPERTIES:  
 CC pH dependence:  
 CC Optimum pH is 4-6;  
 CC -1- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
 CC subunits, the proteolytic products.  
 CC -1- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.  
 CC Secreted, internalised and transferred to late endosomes/lysosomes  
 CC as a proheparanase. In lysosomes, it is processed into the active  
 CC form, the heparanase. The uptake or internalisation of  
 CC proheparanase is mediated by HSPGs. Heparin appears to be a  
 CC competitor and retain proheparanase in the extracellular medium.  
 CC -1- TISSUE SPECIFICITY: Highly expressed in placenta and spleen and  
 CC weakly expressed in lymph node, thymus, peripheral blood  
 CC leukocytes, bone marrow, endothelial cells, fetal liver and tumor

CC tissues.  
 CC -1- PTM: Proteolytically processed. The cleavage of the 65 kDa form  
 CC leads to the generation of a linker peptide, 8 kDa and 50 kDa  
 CC product. The active form, the 8/50 kDa heterodimer, is resistant  
 CC to degradation. Complete removal of the linker peptide appears to  
 CC be a prerequisite to the complete activation of the enzyme.  
 CC -1- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears  
 CC to be essential for its solubility.

Query Match 100.0%; Score 86; DB 1; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TWRHYLLNGRTATR 14  
 Db 294 TWRHYLLNGRTATR 307

## RESULT 2

HPSE\_CHICK  
 ID HPSE\_CHICK STANDARD; PRT; 523 AA.  
 AC Q90YK5;  
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
 DT 01-DEC-2001, sequence version 1.  
 DT 07-FEB-2006, entry version 12.  
 DE Heparanase precursor (EC 3.2.-.-).  
 GN Name=HPSE; Synonyms=HPA;  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA].  
 RX MEDLINE=21369599; PubMed=11387326; DOI=10.1074/jbc.M102462200;  
 RA Goldshmidt O., Zcharia E., Aingorn H., Guatta-Rangini Z., Atzmon R.,  
 RA Michal I., Pecker I., Mitrani E., Vlodavsky I.;  
 RT "Expression pattern and secretion of human and chicken heparanase are  
 RT determined by their signal peptide sequence.";  
 RL J. Biol. Chem. 276:29178-29187(2001).  
 CC -1- FUNCTION: Endoglycosidase which is a cell surface and  
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
 CC proteoglycans (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted  
 CC (By similarity).  
 CC -1- PTM: N-glycosylated (By similarity).  
 CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.

-----  
 Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 Distributed under the Creative Commons Attribution-NonCommercial License

-----  
 EMBL: AY017007; AAK82648.1; -; mRNA.  
 DR Ensembl: ENSGALG00000011203; Gallus gallus.  
 DR InterPro: IPR005199; Glyco\_hydro\_79\_N.  
 DR Pfam: PF03662; Glyco\_hydro\_79n; 1.  
 DR KW Glycoprotein; Hydrolase; Lysosome; Membrane; Signal.  
 FT SIGNAL 1 18 Potential.  
 FT CHAIN 19 523 Heparanase.  
 FT FTID=PRO\_0000042259.  
 FT REGION 137 141 Heparin/HS-binding (By similarity).  
 FT REGION 250 260 Heparin/HS-binding (By similarity).  
 FT ACT\_SITE 204 204 Proton donor (Potential).  
 FT ACT\_SITE 323 323 Nucleophile (Potential).  
 FT CARBOHYD 141 141 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 196 196 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 436 436 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 439 439 N-linked (GlcNAc...) (Potential).  
 SQ SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF81 CRC64;

Query Match 91.9%; Score 79; DB 1; Length 523;  
 Best Local Similarity 85.7%; Pred. No. 0.00064;  
 Matches 12; Conservative 2; Mismatches 0; Indels 0; Gaps 0;



RC STRAIN=C57BL/6J, and MOD; TISSUE=Thymus;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
 RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Chowdhury V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 RA Di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 RA Gustigich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Hill D., Huminecki L., Jacono M., Ikeo K., Iwama A., Ishikawa T.,  
 RA Jakt M., Kanapin A., Katon M., Kawasawa Y., Keiso J., Kitamura H.,  
 RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 RA Mottagui-Tabar S., Mulder N., Nakano N., Nakaguchi H., Ng P.,  
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
 RA Petrovsky N., Piazza S., Reed J.C., Reid J.F., Ring B.Z., Ringwald M.,  
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C., Sheng Y.,  
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 RA Sperling S., Stupka E., Sugtara K., Sultana R., Takenaka Y., Taki K.,  
 RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.N., Tesdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
 RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tagami M., Waki K., Wataniki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome.";  
 RL Science 309:1559-1563(2005).  
 CC -!- FUNCTION: Endoglycosidase which is a cell surface and  
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
 CC proteoglycans. Also implicated in the extravasation of leukocytes  
 CC and tumor cell lines. Contributes to metastasis and angiogenesis  
 CC (By similarity).  
 CC -!- ENZYME REGULATION: Inhibited by EDTA and activated by calcium and  
 CC magnesium (By similarity). Inhibited by laminarin sulfate and, to  
 CC a lower extent, by heparin and sulfamin.  
 CC -!- BIOPHYSICOCHEMICAL PROPERTIES:  
 CC pH dependence:  
 CC Optimum pH is 5;  
 CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
 CC subunits, the proteolytic products.  
 CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.  
 CC Secreted, internalised and transferred to late endosomes/lysosomes  
 CC as a proheparanase. In lysosomes, it is processed into the active  
 CC form, the heparanase. The uptake or internalisation of  
 CC proheparanase is mediated by HSPGs. Heparin appears to be a  
 CC competitor and retain proheparanase in the extracellular medium  
 CC (By similarity).  
 CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form  
 CC leads to the generation of a linker peptide, 8 kDa and 50 kDa  
 CC product. The active form, the 8/50 kDa heterodimer, is resistant  
 CC to degradation. Complete removal of the linker peptide appears to  
 CC be a prerequisite to the complete activation of the enzyme (By  
 CC similarity).  
 CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears  
 CC to be essential for its solubility.  
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms

CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL; AF359507; AAQ15188.1; -; mRNA.  
 CC EMBL; AY077467; AAL76083.1; -; mRNA.  
 CC EMBL; AY151051; AAN41636.1; -; mRNA.  
 CC EMBL; AK040471; BAC30600.1; -; mRNA.  
 CC EMBL; AK154628; BAE32725.1; -; mRNA.  
 CC Ensembl; ENSMUSG00000035273; Mus musculus.  
 CC MGI; MGI:1343124; Hspc.  
 CC GO; GO:0005578; C:extracellular matrix (sensu Metazoa); TAS.  
 CC InterPro; IPR005199; Glyco hydro\_79\_N.  
 CC Pfam; PF03662; Glyco hydro\_79n; 1.  
 CC Calcium; Direct protein sequencing; Glycoprotein; Hydrolase; Lysosome;  
 CC Magnesium; Membrane; Signal. By similarity.  
 CC SIGNAL 1 27  
 CC CHAIN 28 101  
 CC PROPEP 102 149  
 CC CHAIN 150 535  
 CC REGION 150 154  
 CC REGION 262 272  
 CC ACT\_SITE 217 217  
 CC ACT\_SITE 335 335  
 CC CARBOHYD 154 154  
 CC CARBOHYD 192 192  
 CC CARBOHYD 209 209  
 CC CARBOHYD 230 230  
 CC CARBOHYD 451 451  
 CC CONFLICT 206 206  
 CC CONFLICT 212 212  
 CC CONFLICT 230 232  
 CC CONFLICT 335 335  
 CC CONFLICT 342 342  
 CC CONFLICT 455 455  
 CC CONFLICT 531 531  
 CC SEQUENCE 535 AA; 60050 MW; AF19B28B/CD03F7B CRC64;  
 CC Query Match 89.5%; Score 77; DB 1; Length 535;  
 CC Best Local Similarity 85.7%; Pred. No. 0.0014;  
 CC Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 CC  
 CC Qy 1 TWHYYLNGRTATR 14  
 CC |||||  
 CC Db 286 TWHYYLNGRTATK 299  
 CC  
 CC RESULT 5  
 CC Q333X6 SPAJD  
 CC ID Q333X6 SPAJD PRELIMINARY; PRT; 574 AA.  
 CC AC Q333X6;  
 CC DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.  
 CC DT 06-DEC-2005, sequence version 1.  
 CC DT 07-FEB-2006, entry version 3.  
 CC DE Heparanase.  
 CC GN Name=hpa;  
 CC OS Spalax judaei (Blind subterranean mole rat).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 CC OC Muroidae; Spalacidae; Spalacinae; Spalax.  
 CC OX NCBI\_TaxID=134510;  
 CC [1]  
 CC RN NUCLEOTIDE SEQUENCE.  
 CC RP TISSUE=Kidney;  
 CC RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Aviavi A.;  
 CC RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
 CC cloning and identification of a novel splice variant";  
 CC Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).  
 CC [2]  
 CC RN NUCLEOTIDE SEQUENCE.  
 CC RP TISSUE=Kidney;  
 CC RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Aviavi A.;

```
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DR EMBL; AM085493; CAJ30020.1; -; mRNA.
SQ SEQUENCE 574 AA; 64515 MW; 3AEBB13F07451684 CRC64;

Query Match      89.5%; Score 77; DB 2; Length 574;
Best Local Similarity 85.7%; Pred. No. 0.0015;
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
Db 325 TWHYYLNGRIATK 338

RESULT 6
Q333X7_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X7_9RODE PRELIMINARY; PRT; 574 AA.
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax carmeli.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164324;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DR EMBL; AM085492; CAJ30019.1; -; mRNA.
SQ SEQUENCE 574 AA; 64459 MW; 9FD19DCBBD99DE CRC64;

Query Match      89.5%; Score 77; DB 2; Length 574;
Best Local Similarity 85.7%; Pred. No. 0.0015;
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
Db 325 TWHYYLNGRIATK 338

RESULT 7
Q333X8_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X8_9RODE PRELIMINARY; PRT; 574 AA.
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax golani.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=191382;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
```

```
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DR EMBL; AM085491; CAJ30018.1; -; mRNA.
SQ SEQUENCE 574 AA; 64555 MW; 48BEFEC7D0BCB34 CRC64;

Query Match      89.5%; Score 77; DB 2; Length 574;
Best Local Similarity 85.7%; Pred. No. 0.0015;
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
Db 325 TWHYYLNGRIATK 338

RESULT 8
Q333X9_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X9_9RODE PRELIMINARY; PRT; 574 AA.
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax galili.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164323;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DR EMBL; AM085490; CAJ30017.1; -; mRNA.
SQ SEQUENCE 574 AA; 64525 MW; 1635865051B380D0 CRC64;

Query Match      89.5%; Score 77; DB 2; Length 574;
Best Local Similarity 85.7%; Pred. No. 0.0015;
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TWHYYLNGRTATR 14
Db 325 TWHYYLNGRIATK 338

RESULT 9
HPSE_BOVIN
```

```

ID HPSE BOVIN STANDARD; PRT; 545 AA.
AC Q9MYO;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-JUN-2001, sequence version 2.
DT 07-MAR-2006, entry version 15.
DE Heparanase precursor (8C 3.2.-.-) [Contains: Heparanase 8 kDa subunit;
DE Heparanase 50 kDa subunit].
GN Name=HPSE;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RC TISSUE=Placenta;
RX MEDLINE=21176669; PubMed=11277877;
RZ Kizaki K., Nakano H., Nakano H., Takahashi T., Imai K., Hashizume K.;
RT "Expression of heparanase mRNA in bovine placenta during gestation.";
RL Reproduction 121:573-580(2001).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Contributes to metastasis and angiogenesis
CC (By similarity).
CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower
CC extent, by heparin, sulfamin and EDTA. Activated by calcium and
CC magnesium (By similarity).
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products (By similarity).
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes
CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium
CC (By similarity).
CC -!- TISSUE SPECIFICITY: Highly expressed in placenta and weakly in the
CC kidney, lung, spleen and uterus.
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme (By
CC similarity).
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
-----
DR EMBL; AF281160; AAF87301.2; -; mRNA.
DR InterPro; IPR005199; Glyco_hydro_79_N.
DR Pfam; PF03662; Glyco_hydro_79n; 1.
KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;
KW SIGNAL.
FT SIGNAL 1 37 By similarity.
FT CHAIN 38 111 Heparanase 8 kDa subunit (By similarity).
FT FTID=PRO_0000042256.
FT PROPEP 112 159 Linker peptide.
FT FTID=PRO_0000042257.
FT CHAIN 160 545 Heparanase 50 kDa subunit (By
FT similarity).
FT FTID=PRO_0000042258.
FT REGION 160 164 Heparin/HS-binding (Potential).
FT REGION 272 282 Heparin/HS-binding (Potential).
FT ACT_SITE 227 227 Proton donor (Potential).
FT ACT_SITE 345 345 Nucleophile (Potential).
FT CARBOHYD 164 164 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 219 219 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 461 461 N-linked (GlcNAc...) (Potential).

```

---

```

SQ SEQUENCE 545 AA; 61077 MW; FAC4BDFD855B933 CRC64;
Query Match 86.0%; Score 74; DB 1; Length 545;
Best Local Similarity 78.6%; Pred. NO. 0.0042;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 1 TWHYYLNGRTATR 14
Db 296 TWHYYVNGRIATK 309
RESULT 10
QATGCS_TETNG PRELIMINARY; PRT; 255 AA.
AC Q4TGC8;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Chromosome undetermined SCAF3783, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG00001168001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=1549614; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouteau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dessat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Aichmond V., Jubin C., Catelloni V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulin J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McSwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
-----
DR EMBL; CAAB01003783; CAF88054.1; -; Genomic_DNA.
FT NON_TER 1 1
FT NON_TER 255 255
SQ SEQUENCE 255 AA; 28562 MW; 07F542A9C755E3F0 CRC64;
Query Match 76.7%; Score 66; DB 2; Length 255;
Best Local Similarity 76.9%; Pred. NO. 0.036;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 1 TWHYYLNGRTAT 13
Db 130 TWHYYLDGREAS 142
RESULT 11
Q4SYF6_TETNG PRELIMINARY; PRT; 533 AA.
ID Q4SYF6_TETNG

```



Q45TF6;  
AC 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
DT 19-JUL-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Chromosome undetermined SCAF12073, whole genome shotgun sequence.  
DE (Fragment).  
DE  
GN ORNames=GSTENG00010356001;  
GN Tetraodon nigroviridis (Green puffer).  
GN Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
OC Tetraodontidae; Tetraodontidae; Tetraodon.  
OX NCBI\_TaxID=99883;  
OX [1]  
RN NUCLEOTIDE SEQUENCE.  
RP PubMed=15496914; DOI=10.1038/nature03025;  
RX Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,  
RA Mauceli E., Bounneau L., Fischer C., Ozouf-Costaz C., Bernot A.,  
RA Nicaud S., Jaffe D., Fisher S., Lutfailla G., Dossat C., Segurens B.,  
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
RA Anthouard V., Jubin C., Castelli V., Katinka J., Vacherie B.,  
RA Biemont C., Skalli Z., Catrocilo L., Poulin J., De Berardinis V.,  
RA Cruaud C., Duprat S., Brotier P., Coutanceau J.-P., Gouzy J.,  
RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,  
RA Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,  
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.  
RA "Genome duplication in the teleost fish Tetraodon nigroviridis reveals  
RT the early vertebrate proto-karyotype.";  
RN Nature 431:946-957(2004).  
RN [2]  
RN NUCLEOTIDE SEQUENCE.  
RP Genoscope; Whitehead Institute Centre for Genome Research;  
RG Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
RL -! CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; CAAE01012073; CAF94326.1; -; Genomic\_DNA.  
CC  
DR NON\_TER 1  
FT NON\_TER 533  
FT SEQUENCE 533 AA; 60100 MW; 9B00A7C8780100FF CRC64;  
SQ  
  
Query Match 76.7%; Score 66; DB 2; Length 533;  
Best Local Similarity 76.9%; Pred. No. 0.076;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps  
  
Qy 1 TWHYYLNGRTAT 13  
|||:|:|:|:  
Db 250 TWHYYLDGREAS 262  
  
RESULT 12  
ID Q8T108\_BOMMO PRELIMINARY; PRT; 515 AA.  
AC Q8T108;  
DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.  
DT 01-JUN-2002, sequence version 1.  
DT 07-FEB-2006, entry version 10.  
DE Heparanase-like protein.  
DE Names=Bmhpa;  
GN Bombyx mori (Silk moth).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;  
OC Bombycidae; Bombyx.  
OX NCBI\_TaxID=7091;  
OX [1]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=p50; TISSUE=Posterior silkland;  
RC

RA	Koike Y., Mita K., Suzuki M.G., Maeda S., Abe H., Osoegawa K.,
deJong P.J., Shimada T.;	
RT	"Genomic sequence of a 320-kb segment of the Z chromosome of Bombyx
RT	mori containing a kettin ortholog.";
RL	Mol. Genet. Genomics 269:137-149(2003).
CC	-----
CC	Copyrighted by the UniProt Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a>
CC	Distributed under the Creative Commons Attribution-NoDerivs License
CC	-----
DR	ENBL; AB079860; BAB85191.1; -; Genomic DNA.
DR	ENBL; AB090307; BAC10612.1; -; Genomic DNA.
DR	InterPro; IPR005199; Glyco_hydro_79_N.
DR	Pfam; PF03662; Glyco_hydro_79n; 1.
SQ	SEQUENCE 515 AA; 59770 MW; FB8100ABE6EDDADB CRC64;
Query Match	64.0%; Score 55; DB 2; Length 515;
Best Local Similarity	66.7%; Pred. No. 4.1;
Matches	8; Conservative 2; Mismatches 2; Indels 0; Gaps
Qy	1 TWHHYYLNGRTA 12
Db	271 SWHQYYLNSKTA 282
RESULT 13	
Q9PTB2 ICTPU	PRELIMINARY; PRT; 532 AA.
AC	Q9PTB2;
DT	01-MAY-2000, integrated into UniProtKB/TrEMBL.
DT	01-MAY-2000, sequence version 1.
DT	07-FEB-2006, entry version 20.
DE	Tyrosinase.
OS	Itcalurus punctatus (Channell catfish).
OC	Eukaryota; Metazoa; Chordata; Craniata;
OC	Actinopterygii; Neopterygii; Teleostei; Osteiophysi; Siluriformes;
OC	Itcaluridae; Itcalurus.
OX	NCBI_TaxID=7998;
RN	[1]
RP	NUCLEOTIDE SEQUENCE.
TISSUE=Eye;	
RC	Zhang Q., Cooper R.K. Jr., Tiersch T.R.;
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC	-----
CC	Copyrighted by the UniProt Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a>
CC	Distributed under the Creative Commons Attribution-NoDerivs License
CC	-----
DR	ENBL; AF216388; AAF20161.1; -; mRNA.
DR	GO; GO:0016491; F:oxidoreductase activity; IEA.
DR	GO; GO:0005198; F:structural molecule activity; IEA.
DR	GO; GO:0008152; P:metabolism; IEA.
DR	InterPro; IPR002049; EGF laminin.
DR	InterPro; IPR013032; EGF_like_reg.
DR	InterPro; IPR002227; Tyrosinase.
DR	Pfam; PF00053; Laminin EGF; 1.
DR	Pfam; PF00264; Tyrosinase; 1.
DR	PRINTS; PR00092; TYROSINASE.
DR	PROSITE; PS00022; EGF 1; UNKNOWN 1.
DR	PROSITE; PS01248; EGF_LAM 1; UNKNOWN 1.
DR	PROSITE; PS00497; TYROSINASE 1; 1.
DR	PROSITE; PS00498; TYROSINASE 2; 1.
SQ	SEQUENCE 532 AA; 60107 MW; 9BDE703F711CD2EF CRC64;
Query Match	60.5%; Score 52; DB 2; Length 532;
Best Local Similarity	61.5%; Pred. No. 13;
Matches	8; Conservative 3; Mismatches 2; Indels 0; Gaps
Qy	2 WHHHYYLNGRTATR 14
Db	179 WHHHYYLNGRTATR 191

RESULT 14  
O4TB80 TETNG



ID Q4TB80 TETNG PRELIMINARY; PRT; 597 AA.  
AC Q4TB80;  
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
DT 19-JUL-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Chromosome 17 SCAP180, whole genome shotgun sequence. (Fragment).  
GN ORFNames=GSTENG0003868001;  
OS Tetraodon nigroviridis (Green puffer).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
OC Tetraodontidae; Tetraodontidae; Tetraodon.  
OX NCBI\_TaxID=99883;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15496914; DOI=10.1038/nature03025;  
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,  
RA Maudeli E., Bouneau L., Fischer C., Ouzif-Costaz C., Bernot A.,  
RA Nicaut S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,  
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
RA Athouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,  
RA Blemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,  
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,  
RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,  
RA Kallis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,  
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;  
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals  
RT the early vertebrate proto-karyotype.";  
RL Nature 431:946-957(2004).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RG Genoscope; Whitehead Institute Centre for Genome Research;  
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
EMBL; CAAB01007180; CAP89852.1; -; Genomic\_DNA.  
FT NON TER 597 597  
SQ SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;  
  
Query Match 60.5%; Score 52; DB 2; Length 597;  
Best Local Similarity 70.0%; Pred. No. 14;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 TWHHYLYNGR 10  
Db 334 TWQHYWDGR 343  
|||:||||  
-----  
RESULT 15  
OS9462 PYRHO PRELIMINARY; PRT; 1179 AA.  
ID OS9462 PYRHO PRELIMINARY; PRT; 1179 AA.  
AC OS9462;  
DT 01-AUG-1998, integrated into UniProtKB/TrEMBL.  
DT 01-AUG-1998, sequence version 1.  
DT 07-FEB-2006, entry version 30.  
DE 1179aa long hypothetical chromosome assembly protein.  
GN OrderedLocustNames=PH1798;  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=53953;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=OT3;  
RX MEDLINE=98344137; PubMed=9679194; DOI=10.1093/dnares/5.2.55;  
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,

Yamamoto S., Sekine M., Baba S., Kosugi H., Hosoyama A., Nagai Y.,  
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,  
RA Aoki K., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
RA Mauchi Y., Shizuya H., Kikuchi H.;  
RT "Complete sequence and gene organization of the genome of a hyper-  
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
RL DNA Res. 5:55-76(1998).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
EMBL; BA000001; BAA30917.1; -; Genomic\_DNA.  
DR PIR; F71190; F71190.  
DR HSPF; Q9X0R4; 1E69.  
DR SMR; OS9462; 2-167, 1012-1170.  
DR BioCyc; PHOR53953:PH1798-MONOMER; -.  
DR GO; GO:0005694; C:chromosome; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0016887; F:ATPase activity; IEA.  
DR GO; GO:0005525; F:GTP binding; IEA.  
DR GO; GO:0005515; F:protein binding; IEA.  
DR GO; GO:0051277; P:chromosome organization and biogenesis (sen. .; IEA.  
DR InterPro; IPR003439; ABC transporter-like.  
DR InterPro; IPR005289; GTP\_bd.  
DR InterPro; IPR010935; SMC hinge.  
DR InterPro; IPR003395; SMC N.  
DR InterPro; IPR011891; SMC prok A.  
DR InterPro; IPR002017; Spectrin.  
DR Pfam; PF06470; SMC hinge; 1.  
DR Pfam; PF02463; SMC\_N; 2.  
DR ProDom; PD000006; ABC transporter; 1.  
DR TIGRFAMs; TIGR00650; MG442; 1.  
DR TIGRFAMs; TIGR02169; SMC prok A; 1.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 1179 AA; 135657 MW; 24A61BESF3864493 CRC64;  
  
Query Match 59.3%; Score 51; DB 2; Length 1179;  
Best Local Similarity 81.8%; Pred. No. 42;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 4 HYVLYNGRTATR 14  
Db 113 HYVLYNGRTATR 123  
|||:|||||  
-----  
RESULT 16  
Q4WRV8 ASPFU PRELIMINARY; PRT; 124 AA.  
ID Q4WRV8 ASPFU PRELIMINARY; PRT; 124 AA.  
AC Q4WRV8;  
DT 05-JUL-2005, integrated into UniProtKB/TrEMBL.  
DT 05-JUL-2005, sequence version 1.  
DT 07-MAR-2006, entry version 5.  
DE Hypothetical protein.  
GN ORFNames=Afuig14930;  
OS Aspergillus fumigatus (Sartorya fumigata).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5085;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RX STRAIN=Af293 / CBS 101355 / FGSC A1100;  
RX PubMed=16372009; DOI=10.1038/nature04332;  
RA Nierman W.C., Pain A., Anderson M.J., Wortman J.R., Kim H.S.,  
RA Arroyo J., Chen D., Collins K., Archer D.B., Bermejo C., Bennett J.W.,  
RA Bowyer P., Berran D., Collins N., Coulson R., Davies K., Dyer P.S.,  
RA Farman M., Fedorova N., Fedorova N.D., Feldblyum T.V., Fischer R.,  
RA Fosker N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,  
RA Goldman G.H., Gomi K., Griffith-Jones S., Gwilliam R., Haas B.J.,  
RA Haas H., Harris D.E., Horiuchi H., Huang J., Humphray S., Jimenez J.,  
RA Keller N., Khouri H., Kitamoto K., Kobayashi T., Konzack S.,  
RA Kulkarni R., Kumagai T., Laiton A., Latge J.-P., Li W., Lord A.,  
RA Lu C., Majoros W.H., May G.S., Miller B.L., Mohamoud Y., Molina M.,

RA Monod M., Mouyna I., Mulligan S., Murphy L.D., O'Neil S., Paulsen I.,  
RA Penalba M.A., Perteu C., Pritchard B.L., Quail M.A.,  
RA Rabinowitz E., Rawlins N., Rajandream M.A., Reichard U.,  
RA Renauld H., Robson G.D., Rodriguez de Cordoba S., Rodriguez-Pena J.M.,  
RA Ronning C.M., Rutter S., Salzberg S.L., Sanchez M.,  
RA Sanchez-Ferreiro J.C., Saunders D., Seeger K., Squares R., Squares S.,  
RA Takeuchi M., Tekala F., Turner G., Vazquez de Aldana C.R., Weidman J.,  
RA White O., Woodward J.R., Yu J.-H., Fraser C.M., Galagan J.E., Asai K.,  
RA Machida M., Hall N., Barrell B.G., Denning D.W.;  
RT "Genomic sequence of the pathogenic and allergenic filamentous fungus  
RT *Aspergillus fumigatus*.";  
RL Nature 438:1151-1156(2005).  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AAHF01000004; EAL90824.1; -; Genomic\_DNA.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 124 AA; 14143 MW; DF32481969DBD831 CRC64;  
  
Query Match 58.1%; Score 50; DB 2; Length 124;  
Best Local Similarity 57.1%; Pred. No. 6.1;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
  
QY 1 TWHHVYLNQRTATR 14  
||||| :  
DB 64 TWHHGYLGGQVRER 77  
  
RESULT 17  
ID Q2KXXB BORAV PRELIMINARY; PRT; 52 AA.  
AC Q2KXXB;  
DT 07-MAR-2006, integrated into UniProtKB/TrEMBL.  
DT 07-MAR-2006, sequence version 1.  
DT 07-MAR-2006, entry version 1.  
DE Hypothetical protein.  
GN ORFNames=BAV2400;  
OS Bordetella avium 197N.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Alcaligenaceae; Bordetella.  
OX NCBI\_TaxID=360910;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=197N;  
RA Sebaihia M.;  
RT "The genome sequence of the poultry pathogen *Bordetella avium*, and  
RT genomic comparisons with related species infecting mammals.";  
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AM167904; CAJ50010.1; -; Genomic\_DNA.  
KW Hypothetical protein.  
SQ SEQUENCE 52 AA; 6063 MW; 2C7181A0BCAACAA3 CRC64;  
  
Query Match 57.0%; Score 49; DB 2; Length 52;  
Best Local Similarity 53.8%; Pred. No. 3.6;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 WHHYVYLNQRTATR 14  
||||| :  
DB 13 WHHHWLNTRVNTK 25  
  
RESULT 18  
ID Q7PIC9 CHRVO PRELIMINARY; PRT; 138 AA.  
AC Q7PIC9;

DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.  
DT 15-DEC-2003, sequence version 1.  
DT 07-FEB-2006, entry version 10.  
DE Hypothetical protein.  
GN OrderedLocusNames=CV0284; ORFNames=CV\_0284;  
OS Chromobacterium violaceum.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
OC Neisseriaceae; Chromobacterium.  
OX NCBI\_TaxID=536;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=ATCC 12472 / DSM 30191;  
RX MEDLINE=22882880; PubMed=14500782; DOI=10.1073/pnas.1832124100;  
RA Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,  
RA Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,  
RA Alves-Gomes J.A., Andrade E.M., Aratipe J., de Azaújo M.F.F.,  
RA Astolfi-Filho S., Azevedo V., Baptista A.J., Bataus L.A.M.,  
RA Batista J.S., Beio A., van den Berg C., Bogo M., Bonatto S.,  
RA Bordignon J., Brígido M.M., Brito C.A., Brocchi M., Burity H.A.,  
RA Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,  
RA Carvalho C.M.B., Cascado J.C.M., Cavada B.S., Chueire L.M.O.,  
RA Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcão C.L.,  
RA Fantinatti F., Farias I.P., Felipe M.S.S., Ferrari L.P., Ferro J.A.,  
RA Ferro M.I.T., Franco G.R., Freitas N.S.A., Furlan L.R., T.B.,  
RA Gazzinelli R.T., Gomes E.A., Gonçalves P.R., Grangeiro T.B.,  
RA Grattapaglia D., Grisard E.C., Hanna E.S., Jardim S.N., Laurino J.,  
RA Leclercq T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,  
RA Madeira H.M.F., Manfio G.P., Maranhão A.O., Martins W.S.,  
RA di Mauro S.M.Z., de Medeiros S.R.B., Meissner R.V., Moreira M.A.M.,  
RA Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,  
RA Paixão R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,  
RA Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Petrich D.P.,  
RA Ramalho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,  
RA Santos E.B.P., Santos F.R., Schneider M.P.C., Seunarez H.N.,  
RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,  
RA Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,  
RA Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Urmenyi T.,  
RA Vettore A., Wassem R., Zaha A., Simpson A.J.G.;  
RT "The complete genome sequence of *Chromobacterium violaceum* reveals  
RT remarkable and exploitable bacterial adaptability.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AE016825; AAQ57963.1; -; Genomic\_DNA.  
DR Biocyc; CV10243365:CV0284-MONOMER; -;  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 138 AA; 15009 MW; BEC7720F3E6500CF CRC64;  
  
Query Match 54.7%; Score 47; DB 2; Length 138;  
Best Local Similarity 70.0%; Pred. No. 20;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 WHHYVYLNQRT 11  
||||| :  
DB 45 WHHCYVHGRT 54  
  
RESULT 19  
ID Q5C4P3 SCHJA PRELIMINARY; PRT; 215 AA.  
AC Q5C4P3;  
DT 12-APR-2005, integrated into UniProtKB/TrEMBL.  
DT 08-NOV-2005, sequence version 2.  
DT 07-FEB-2006, entry version 6.  
DE SJCHGC07775 protein (Fragment).  
OS Schistosoma japonicum (Blood fluke).  
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;  
OC Schistosomatoidea; Schistosomatidae; Schistosoma.  
OX NCBI\_TaxID=6182;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.

RA Liu F., Lu J., Hu W., Wang S.-Y., Cui S.-J., Chi M., Yan Q.,  
 RA Wang X.-R., Song H.-D., Xu X.-N., Wang J.-J., Zhang X.-L., Wang Z.-Q.,  
 RA Xue C.-L., Brindley P.J., McManus D.P., Yang P.-Y., Feng Z., Chen Z.,  
 RA Han Z.-G.;  
 RT "New Perspectives on Host-parasite Interplay by Comparative  
 RT Transcriptomic and Proteomic Analyses of the Human Blood Fluke,  
 RT *Schistosoma japonicum*,"  
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 DR EMBL; AY809492; AAX25381.2; -; mRNA.  
 FT NON\_TER 1 215  
 FT NON\_TER 215 215  
 SQ SEQUENCE 215 AA; 25120 MW; D02C316C7C2B80C3 CRC64;  
 Query Match 54.7%; Score 47; DB 2; Length 215;  
 Best Local Similarity 63.6%; Pred. No. 32;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 WHYYLINGRTA 12  
 DB 85 WHPYLSNRSTS 95  
 RESULT 20  
 Q7VJ45 HELHP PRELIMINARY; PRT; 382 AA.  
 AC Q7VJ45;  
 DT 01-OCT-2003, integrated into UniProtKB/TrEMBL.  
 DT 01-OCT-2003, sequence version 1.  
 DT 07-FEB-2006, entry version 17.  
 DE Citrate synthase PrpC (EC 2.3.3.1).  
 GN Names:prpC; OrderedLocusNames:HH0398; ORFNames:HH\_0398;  
 OS Helicobacter hepaticus.  
 GN Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
 OC Helicobacteraceae; Helicobacter.  
 OX NCBI\_TaxID=32025;  
 [1]  
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=ATCC 51449 / 381;  
 RA MEDLINE=22709201; PubMed=12810954; DOI=10.1073/pnas.1332093100;  
 RA Suerbaum S., Jøsenhans C., Steerzenbach T., Drescher B., Brandt P.,  
 RA Bell M., Droege M., Fartmann B., Fischer H.-P., Ge Z., Hoerster A.,  
 RA Holland R., Klein K., König J., Macko L., Mendz G.L., Nyakatura G.,  
 RA Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;  
 RT "The complete genome sequence of the carcinogenic bacterium  
 RT *Helicobacter hepaticus*,"  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906 (2003).  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 DR EMBL; AE017125; AAP76995.1; -; Genomic\_DNA.  
 DR HSSP; O34002; IA59.  
 DR BioCyc; HHPF235279:HH0398-MONOMER; -.  
 DR GO; GO:0005737; Cytoplasm; IEA.  
 DR GO; GO:0008415; F:acyltransferase activity; IEA.  
 DR GO; GO:0004108; F:citrate (5S)-synthase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006092; P:main pathways of carbohydrate metabolism; IEA.  
 DR InterPro; IPR011278; Cit\_synth\_II.  
 DR InterPro; IPR002020; Citrate synth.  
 DR PANTHER; PTHR11739; Citrate synth; 1.  
 DR Pfam; PF00285; Citrate synth; 1.  
 DR PRINTS; PRO0143; CITRYSNTHASE.  
 DR TIGRFAMs; TIGR01800; cit\_synth\_II; 1.  
 DR Acyltransferase; Complete proteome; Transferase.  
 SQ SEQUENCE 382 AA; 42813 MW; C51D9727E13656D3 CRC64;  
 Query Match 54.7%; Score 47; DB 2; Length 382;  
 Best Local Similarity 66.7%; Pred. No. 57;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 WHYYLINGR 10  
 DB 143 WHYYHHNGK 151  
 RESULT 21  
 Q41029 GIBZE PRELIMINARY; PRT; 383 AA.  
 ID Q41029 GIBZE  
 AC Q41029;  
 DT 16-AUG-2005, integrated into UniProtKB/TrEMBL.  
 DT 16-AUG-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 4.  
 DE Hypothetical protein.  
 GN ORFNames=FG09109.1;  
 OS Gibberella zeae (Fusarium graminearum).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Eurotiomycetidae; Hypocreales; Nectriaceae; Gibberella.  
 OX NCBI\_TaxID=5518;  
 RN [1]  
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=PH-1 / NRRL 31084;  
 RA Birren B.W., Nusbaum C., Abouelleil A., Allen N., Anderson S.,  
 RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,  
 RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,  
 RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., Dearliano K.,  
 RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,  
 RA Erickson J., Faro S., Ferreira P., Fitzgerald M., Gage D.,  
 RA Galagan J.E., Gardyna S., Gnerre S., Graham L., Grand-Pierre N.,  
 RA Hafez N., Hagopian D., Hagos B., Hall J., Horton L., Hulme W.,  
 RA Iliev I., Jaffe D., Johnson R., Jones C., Kamal M., Kamat A.,  
 RA Karatas A., Keile C., Landers T., Levine R., Lindblad-Toh K., Liu G.,  
 RA Lui A., Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J.,  
 RA Manning J., Matthews C., Mauceli E., McCarthy M., Meldrim J.,  
 RA Menues L., Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C.,  
 RA Nicol R., Nielsen C.B., Norbu C., O'Connor T., O'Donnell P.,  
 RA O'Neill D., Oliver J., Peterson K., Phunkhang P., Pierre N.,  
 RA Purcell S., Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C.,  
 RA Rogov P., Roman J., Schauer S., Schupbach R., Seaman S., Severy P.,  
 RA Smirnov S., Smith C., Spencer B., Stange-Thomann N., Stojanovic N.,  
 RA Stubbs M., Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,  
 RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,  
 RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,  
 RA Lander E.S.;  
 RT "Fusarium graminearum genome sequence,"  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 DR EMBL; AACM01000370; EAA78159.1; -; Genomic\_DNA.  
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 383 AA; 42750 MW; 993AC5392560831A CRC64;  
 Query Match 54.7%; Score 47; DB 2; Length 383;  
 Best Local Similarity 72.7%; Pred. No. 57;  
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 3 HHYYLINGRTAT 13  
 DB 272 HHYYFAGRPAT 282  
 RESULT 22  
 KCMA1 CANFA STANDARD; PRT; 1159 AA.  
 ID KCMA1 CANFA  
 AC Q28265;

DT 13-APR-2004, integrated into UniProtKB/Swiss-Prot.  
DT 13-APR-2004, sequence version 2.  
DT 07-MAR-2006, entry version 40.  
DE Calcium-activated potassium channel alpha subunit 1 (Calcium-activated  
DE potassium channel, subfamily M, alpha subunit 1) (Maxi K channel)  
DE (MaxiK) (BK channel) (K(VCA)alpha) (BKCa alpha) (KCa1.1) (Slowpoke  
DE homolog) (Slo homolog) (Slo-alpha) (Slo1) (Fragment).  
GN Name=KCNA1; Synonyms=KCNA;  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
OC Canis.  
OX NCBI\_TaxID=9615;  
[1]  
RN NUCLEOTIDE SEQUENCE [MRNA], FUNCTION, AND TISSUE SPECIFICITY.  
RC TISSUE=Colon smooth muscle;  
RX MEDLINE=97053370; PubMed=8897882;  
RA Vogalis F., Vincent T., Qureshi I., Schmalz F.M., Ward M.W.,  
RA Sanders K.M., Horowitz B.;  
RT "Cloning and expression of the large-conductance Ca(2+)-activated K+  
RT channel from colonic smooth muscle.";  
RL Am. J. Physiol. 271:G629-G639(1996).  
CC -!- FUNCTION: Potassium channel activated by both membrane  
CC depolarization or increase in cytosolic Ca(2+) that mediates  
CC export of K(+). It is also activated by the concentration of  
CC cytosolic Mg(2+). Its activation dampens the excitatory events  
CC that elevate the cytosolic Ca(2+) concentration and/or depolarize  
CC the cell membrane. It therefore contributes to repolarization of  
CC the membrane potential. Plays a key role in controlling of  
CC excitability in a number of systems, such as regulation of the  
CC contraction of smooth muscle, the tuning of hair cells in the  
CC cochlea, regulation of transmitter release, and innate immunity.  
CC In smooth muscles, its activation by high level of Ca(2+), caused  
CC by ryanodine receptors in the sarcoplasmic reticulum, regulates  
CC the membrane potential. In cochlea cells, its number and kinetic  
CC properties partly determine the characteristic frequency of each  
CC hair cell and thereby helps to establish a tonotopic map. Kinetics  
CC of KCNA1 channels are determined by alternative splicing,  
CC phosphorylation status and its combination with modulating beta  
CC subunits. Highly sensitive to both ibertoxin (IbTx) and  
CC charybotoxin (CTX) (By similarity).  
CC -!- SUBUNIT: Homotrimer; which constitutes the calcium-activated  
CC potassium channel. Interacts with beta subunits KCNB1, KCNB2,  
CC KCNB3 and KCNB4. Beta subunits are accessory, and modulate its  
CC activity (By similarity).  
CC -!- SUBCELLULAR LOCATION: Membrane; multi-pass membrane protein (By  
CC similarity).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=1;  
CC Comment=May be partially controlled by hormonal stress. A number  
CC of isoforms are produced;  
CC Name=1;  
CC IsoId=Q28265-1; Sequence=Displayed;  
CC -!- TISSUE SPECIFICITY: Expressed in all vascular and smooth muscles.  
CC -!- DOMAIN: The S0 segment is essential for the modulation by the  
CC accessory beta subunits KCNB1, KCNB2, KCNB3 and KCNB4 (By  
CC similarity).  
CC -!- DOMAIN: The S4 segment, which is characterized by a series of  
CC positively charged amino acids at every third position, is part  
CC of the voltage-sensor (By similarity).  
CC -!- DOMAIN: The pore-forming domain (also referred as P region) is  
CC imbedded into the membrane, and forms the selectivity filter of  
CC the pore. It contains the signature sequence of potassium channels  
CC that displays selectivity to potassium (By similarity).  
CC -!- DOMAIN: The BK-T1 domain mediates the homotrimerization, thereby  
CC promoting the assembly of monomers into functional potassium  
CC channel. It includes the RCK domain, which contains binding sites  
CC for Ca(2+) and Mg(2+) (By similarity).  
CC -!- DOMAIN: The calcium bowl constitutes one of the Ca(2+) sensors and  
CC probably acts as a Ca(2+)-binding site. There are however other  
CC Ca(2+) sensors regions required for activation of the channel (By  
CC similarity).  
CC -!- PTM: Phosphorylated (Probable). Phosphorylation by kinases such as

CC PKA and/or PKG. In smooth muscles, phosphorylation affects its  
CC activity (By similarity).  
CC -!- MISCELLANEOUS: The protein was initially thought to contain two  
CC functionally distinct parts: The core channel (from the N-terminus  
CC to the S9 segment) that mediates the channel activity, and the  
CC cytoplasmic tail (from the S9 segment to the C-terminus) that  
CC mediates the calcium sensing. The situation is however more  
CC complex, since the core channel contains binding sites for Ca(2+)  
CC and Mg(2+).  
CC -!- SIMILARITY: Belongs to the potassium channel family. Calcium-  
CC activated subfamily.  
CC -!- SIMILARITY: Contains 1 BK-T1 domain.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL: U41001; AAA84000.1; -; mRNA.  
CC HSP; O27564; LNQ.  
CC Ensembl: ENSCAFG0000015465; Canis familiaris.  
CC InterPro: IPR003929; BK channel\_alpha.  
CC InterPro: IPR005821; Ion trans.  
CC InterPro: IPR001622; K+channel\_pore.  
CC InterPro: IPR005820; M+channel\_hlg.  
CC Pfam: PF03493; BK\_channel\_a; 1.  
CC Pfam: PF00520; Ion\_trans; 1.  
CC PRINTS: PR01449; BKCHANNELA.  
CC KW Alternative splicing; Calcium; Ion transport; Ionic channel;  
KW Magnesium; Membrane; Metal-binding; Phosphorylation; Potassium;  
KW Potassium channel; Potassium transport; Transmembrane; Transport;  
KW Voltage-gated channel.  
CC CHAIN <1 1159 Calcium-activated potassium channel alpha  
FT subunit 1.  
FT /FTID=PRO\_0000054131.  
FT TOPO\_DOM <1 24 Extracellular (Potential).  
FT TRANSEMEM 25 45 Segment S0 (Potential).  
FT TOPO\_DOM 46 116 Cytoplasmic (Potential).  
FT TRANSEMEM 117 137 Segment S1 (Potential).  
FT TOPO\_DOM 138 152 Extracellular (Potential).  
FT TRANSEMEM 153 173 Segment S2 (Potential).  
FT TOPO\_DOM 174 177 Cytoplasmic (Potential).  
FT TRANSEMEM 178 198 Segment S3 (Potential).  
FT TOPO\_DOM 199 202 Extracellular (Potential).  
FT TRANSEMEM 203 223 Segment S4 (Potential).  
FT TOPO\_DOM 224 238 Cytoplasmic (Potential).  
FT TRANSEMEM 239 259 Segment S5 (Potential).  
FT TRANSEMEM 306 326 Segment S6 (Potential).  
FT TOPO\_DOM 327 1159 Cytoplasmic (Potential).  
FT DOMAIN 327 493 BK-T1.  
FT REGION 274 296 Pore-forming, in membrane (Potential).  
FT REGION 341 430 RCK.  
FT REGION 494 514 Segment S7.  
FT REGION 551 571 Segment S8.  
FT REGION 717 737 Segment S9.  
FT REGION 912 932 Segment S10.  
FT MOTIF 290 293 Selectivity for potassium.  
FT MOTIF 883 905 Calcium bowl.  
FT METAL 377 377 Magnesium (By similarity).  
FT METAL 400 400 Magnesium (By similarity).  
FT METAL 402 402 Magnesium (By similarity).  
FT MOD\_RES 643 643 Phosphothreonine (By similarity).  
FT MOD\_RES 645 645 Phosphoserine (By similarity).  
FT MOD\_RES 857 857 Phosphoserine (By similarity).  
FT MOD\_RES 858 858 Phosphoserine (By similarity).  
FT NON\_TER 1 1  
SQ SEQUENCE 1159 AA; 130296 MW; D8CF38C32971BB64 CRC64;  
Query Match 54.7%; Score 47; DB 1; Length 1159;  
Best Local Similarity 77.8%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Oy 1 TWHYYLNG 9  
Db 526 TWHYYLNG 534

RESULT 23  
 Q8BSE9\_MOUSE PRELIMINARY; PRT; 64 AA.  
 AC Q8BSE9;  
 DT 01-MAR-2003, sequence version 1.  
 DT 07-FEB-2006, entry version 18.  
 DE 12 days embryo embryonic body between diaphragm region and neck cDNA,  
 DE product: sialyltransferase 5, full insert sequence. (Fragment).  
 GN Name: St3gal2; Synonyms: Siat5;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and  
 RC neck;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning."  
 RL Methods Enzymol. 303:19-44 (1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and  
 RC neck;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
 RA Ambesi-Impombato A., Apeeler R., Aturaliya R.N., Bailey T.L.,  
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dall'Amico B.P., de Bono B., Della Gatta G.,  
 RA di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Hill D., Hummel L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
 RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 RA Mottagui-Tabar S., Mulder N., Nakano N., Nakaguchi H., Ng P.,  
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
 RA Petrovsky N., Piazza S., Reed J.C., Pavan W.J., Ring B.Z., Ringwald M.,  
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 RA Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,  
 RA Tammojä K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Inamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
 RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome."  
 RL Science 309:1559-1563 (2005).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and

RC neck;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RG RIKEN Genome Exploration Research Group, and Genome Science Group  
 RT (Genome Network Core Team) and the FANTOM Consortium;  
 RL "Antisense Transcription in the Mammalian Transcriptome."  
 RL Science 309:1564-1566 (2005).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and  
 RC neck;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogi A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia K., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer J.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.A.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi L., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayaseu N.,  
 RA Hironaka-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yaunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573 (2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and  
 RC neck;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischnann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690 (2001).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and  
 RC neck;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayaseu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to

RT prepare full-length cDNA libraries for rapid discovery of new genes." ;  
 RL Genome Res. 10:1617-1630(2000).

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;

RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;

RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,

RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,

RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,

RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,

RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,

RA Yoneda Y., Ishikawa T., Ozawa Y., Tanaka T., Matsuura S., Kawai J.,

RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y. ;

RT "RIKEN integrated sequence analysis (RISA) system-384-format

RT sequencing pipeline with 384 multicapillary sequencer." ;

RL Genome Res. 10:1757-1771(2000).

RN NUCLEOTIDE SEQUENCE.

RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;

RC Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,

RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,

RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,

RA Hori F., Inotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,

RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,

RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,

RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,

RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,

RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tegami M.,

RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,

RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y. ;

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

DR EMBL; AK034554; BAC28752.1; -; mRNA.

DR Ensembl; ENSMUSG00000031749; Mus musculus.

DR MGI; MGI:99427; Slats.

DR MGI; MGI:99427; Slats.

DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.

KW Glycosyltransferase; Transferase.

FT NON\_TER 1 1

SQ SEQUENCE 64 AA; 7426 MW; B7CAB629D166DD6E CRC64;

Query Match 53.5%; Score 46; DB 2; Length 64;

Best Local Similarity 63.6%; Pred. No. 13;

Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHYYLNGRTA 12

DB 21 WHYYENNRYA 31

RESULT 24

RA Bouhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,

RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., DeArellano K.,

RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,

RA Erickson J., Faro S., Ferreira P., Fitzgerald M., Gage D.,

RA Galagan J.E., Gardyna S., Gnerre S., Graham L., Grand-Pierre N.,

RA Hafez N., Hegopian D., Hagos B., Hall J., Horton L., Hulme W.,

RA Iliev I., Jaffe D., Johnson R., Jones C., Kamal M., Kamat A.,

RA Karatas A., Kellis C., Landers T., Levine R., Lindblad-Toh K., Liu G.,

RA Lui A., Ma L.-J., Mabbitt R., MacLean C., Macdonald P., Major J.,

RA Manning J., Matthews C., Mauceli E., McCarthy M., Meldrim J.,

RA Menes L., Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C.,

RA Nicol R., Nielsen C.B., Norbu C., O'Connor T., O'Donnell P.,

RA O'Neil D., Oliver J., Peterson K., Phunkhang P., Pierre N.,

RA Purcell S., Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C.,

RA Rogov P., Roman J., Schauer S., Schupbach R., Seaman S., Severy P.,

RA Smirnov S., Smith C., Spencer B., Stange-Thomann N., Stojanovic N.,

RA Stubbs M., Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,

RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,

RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,

RA Lander E.S. ;

RT "Fusarium graminearum genome sequence." ;

RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.

CC -!- CAUTION: The sequence shown here is derived from an

CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is

CC preliminary data.

CC -----

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

DR EMBL; AAC01000256; EAA73927.1; -; Genomic DNA.

DR GO; GO:0008080; F:N-acetyltransferase activity; IEA.

KW Complete proteome; Hypothetical protein.

SQ SEQUENCE 186 AA; 21114 MW; E758A99EC024AB4D CRC64;

Query Match 53.5%; Score 46; DB 2; Length 186;

Best Local Similarity 66.7%; Pred. No. 40;

Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 HHYYLNGRTATR 14

DB 153 HHYYLNGNSAYR 164

RESULT 25

Q86ZK9\_PODAN PRELIMINARY; PRT; 190 AA.

ID Q86ZK9\_PODAN PRELIMINARY; PRT; 190 AA.

AC Q86ZK9; integrated into UniProtKB/TrEMBL.

DT 01-JUN-2003, sequence version 1.

DT 07-FEB-2006, entry version 8.

DE Similar to N-Acetyltransferase.

OS Podospora anserina.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariomycetidae; Sordariales; Lasiosphaeriaceae; Podospora.

OX NCBI\_TaxID=5145;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RA Genoscope;

RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

DR EMBL; BX088700; CAD60715.1; -; Genomic DNA.

DR GO; GO:0008080; F:N-acetyltransferase activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR InterPro; IPR000182; GCN5acetyl trans.

DR Pfam; PF00583; Acetyltransf\_1; I.

DR Transferase.

KW SEQUENCE 190 AA; 21849 MW; 98B095A23BFEF133 CRC64;

Query Match 53.5%; Score 46; DB 2; Length 190;

Best Local Similarity 66.7%; Pred. No. 41;



```

Matches      8;  Conservative      1;  Mismatches      3;  Indels      0;  Gaps      0;

QY      3 HHYLYNGRTATR 14
|      ||||| :|
Db      154 HRYLYNGNSAYR 165

RESULT 26
QZUL82 ASPOR
ID      QZUL82 ASPOR      PRELIMINARY;      PRT;      208 AA.
AC      QZUL82;
DT      24-JAN-2006, integrated into UniProtKB/TrEMBL.
DT      24-JAN-2006, sequence version 1.
DT      07-MAR-2006, entry version 3.
DE      N-acetyltransferase.
GN      ORFNames=AO090003000513;
OS      Aspergillus oryzae.
OC      Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC      Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX      NCBI_TaxID=5062;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=RTB 40;
RX      PubMed=16372010; DOI=10.1038/nature04300;
RA      Machida M., Asai K., Sano M., Tanaka T., Kumagai T., Terai G.,
RA      Kusumoto K., Arima T., Akita O., Kashiwagi Y., Abe K., Gomi K.,
RA      Horiuchi H., Kitamoto K., Kobayashi T., Takeuchi M., Denning D.W.,
RA      Galagan J.E., Nierman W.C., Yu J., Archer D.B., Bennett J.W.,
RA      Bhatnagar D., Cleveland T.E., Fedorova N.D., Gotoh O., Horikawa H.,
RA      Hosoyama A., Ichinomiya M., Igarashi R., Iwashita K., Juvvadi P.R.,
RA      Kato M., Kato Y., Kin T., Kokubun A., Maeda H., Maeyama N.,
RA      Maruyama J., Nagasaki H., Nakajima T., Oda K., Okada K., Paulsen I.,
RA      Sakamoto K., Sawano T., Takahashi M., Takase K., Terabayashi Y.,
RA      Wortman J.R., Yamada O., Yamagata Y., Anazawa H., Hata Y., Koide Y.,
RA      Komori T., Koyana Y., Minetoki T., Suharnan S., Tanaka A., Isono K.,
RA      Kuhara S., Ogasawara N., Kikuchi H.;
RT      "Genome sequencing and analysis of Aspergillus oryzae.";
RL      Nature 438:1157-1161(2005).
CC      Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC      Distributed under the Creative Commons Attribution-NoDerivs License
CC      EMBL; AP007155; BAB57683.1; -; Genomic DNA.
DR      GO; GO:0016740; F:transferase activity; IEA.
KW      Transferase.
SQ      SEQUENCE      208 AA;      23330 MW;      EEB1E4BFA8F469F6 CRC64;

Query Match      53.5%;      Score 46;      DB 2;      Length 208;
Best Local Similarity      66.7%;      Pred. No. 45;
Matches      8;      Conservative      1;      Mismatches      3;      Indels      0;      Gaps      0;

QY      3 HHYLYNGRTATR 14
|      ||||| :|
Db      155 HRYLYNGNSAYR 166

RESULT 27
Q7S4E2 NEUCR
ID      Q7S4E2 NEUCR      PRELIMINARY;      PRT;      213 AA.
AC      Q7S4E2;
DT      15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT      15-DEC-2003, sequence version 1.
DT      07-FEB-2006, entry version 10.
DE      Hypothetical protein.
GN      ORFNames=NCU02417.1;
OS      Neurospora crassa.
OC      Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC      Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX      NCBI_TaxID=5141;
RN      [1]
RP      NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC      STRAIN=74-OR23-1A / FGSC 987;
RL      Nature 438:1105-1115(2005).
CC      Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC      Distributed under the Creative Commons Attribution-NoDerivs License
CC      EMBL; APO07155; BAB57683.1; -; Genomic DNA.
DR      GO; GO:0016740; F:transferase activity; IEA.
KW      Transferase.
SQ      SEQUENCE      208 AA;      23330 MW;      EEB1E4BFA8F469F6 CRC64;

Query Match      53.5%;      Score 46;      DB 2;      Length 208;
Best Local Similarity      66.7%;      Pred. No. 45;
Matches      8;      Conservative      1;      Mismatches      3;      Indels      0;      Gaps      0;

QY      3 HHYLYNGRTATR 14
|      ||||| :|
Db      155 HRYLYNGNSAYR 166

RESULT 27
Q7S4E2 NEUCR
ID      Q7S4E2 NEUCR      PRELIMINARY;      PRT;      213 AA.
AC      Q7S4E2;
DT      15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT      15-DEC-2003, sequence version 1.
DT      07-FEB-2006, entry version 10.
DE      Hypothetical protein.
GN      ORFNames=NCU02417.1;
OS      Neurospora crassa.
OC      Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC      Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX      NCBI_TaxID=5141;
RN      [1]
RP      NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC      STRAIN=74-OR23-1A / FGSC 987;
RL      Nature 438:1105-1115(2005).
CC      Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC      Distributed under the Creative Commons Attribution-NoDerivs License
CC      EMBL; APO07155; BAB57683.1; -; Genomic DNA.
DR      GO; GO:0016740; F:transferase activity; IEA.
KW      Transferase.
SQ      SEQUENCE      208 AA;      23330 MW;      EEB1E4BFA8F469F6 CRC64;

Query Match      53.5%;      Score 46;      DB 2;      Length 213;
Best Local Similarity      66.7%;      Pred. No. 46;
Matches      8;      Conservative      1;      Mismatches      3;      Indels      0;      Gaps      0;

QY      3 HHYLYNGRTATR 14
|      ||||| :|
Db      176 HRYLYNGNSAYR 187

RESULT 28
Q5B387 EMENI
ID      Q5B387 EMENI      PRELIMINARY;      PRT;      213 AA.
AC      Q5B387;
DT      26-APR-2005, integrated into UniProtKB/TrEMBL.
DT      26-APR-2005, sequence version 1.
DT      07-MAR-2006, entry version 6.
DE      Hypothetical protein.
GN      ORFNames=AN4993.2;
OS      Aspergillus nidulans FGSC A4.
OC      Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC      Eurotiales; Trichocomaceae; Emericella.
OX      NCBI_TaxID=227321;
RN      [1]
RP      NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC      STRAIN=FGSC 4;
RX      PubMed=16372000; DOI=10.1038/nature04341;
RA      Galagan J.E., Calvo S.E., Cuomo C., Ma L.-J., Wortman J.R.,
RA      Batzoglou S., Lee S.-I., Bastuerkmen M., Spevak C.C., Clutterbuck J.,
RA      Kapitonov V., Jurka J., Scaccocchio C., Farman M., Butler J.,
RA      Purcell S., Harris S., Braus G.H., Draht O., Busch S., D'Enfert C.,
RA      Bouchier C., Goldman G.H., Bell-Pedersen D., Griffiths-Jones S.,
RA      Doonan J.H., Yu J., Vienken K., Pain A., Freitag M., Selker E.U.,
RA      Archer D.B., Penalva M.A., Oakley B.R., Momany M., Tanaka T.,
RA      Kumagai T., Asai K., Machida M., Nierman W.C., Denning D.W.,
RA      Caddick M., Hynes M., Paolletti M., Fischer R., Miller B.L., Dyer P.S.,
RA      Sachs M.S., Osmani S.A., Birren B.W.;
RT      "Sequencing of Aspergillus nidulans and comparative analysis with A.
RT      fumigatus and A. oryzae.";
RL      Nature 438:1105-1115(2005).
CC      Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC      Distributed under the Creative Commons Attribution-NoDerivs License
CC      EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC      preliminary data.

```

```

RA      Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA      Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA      Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA      Qui D., Ianakiev P., Bell-Pedersen D., Nelson M.A.,
RA      Werner-Washburne M., Selitrennikoff C.P., Kinsey J.A., Braun E.L.,
RA      Zelter A., Schulte U., Kothe G.O., Jedd G., Mewes H.-W., Staben C.,
RA      Marcotte E., Greenberg D., Roy A., Foley K., Naylor J.,
RA      Stange-Thomann N., Barrett R., Gnerre S., Kamal M., Kamvysselis M.,
RA      Mauceli E., Biele C., Rudd S., Frisman D., Krystofova S.,
RA      Raasmussen C., Metznerberg R.L., Perkins D.D., Kroken S., Cogoni C.,
RA      Macino G., Catchside D.E.A., Li W., Pratt R.J., Osmani S.A.,
RA      DeSouza C.P.C., Glass N.L., Orbach M.J., Berglund J.A., Voelker R.,
RA      Yarden O., Plamann M., Seiler S., Dunlap J.C., Radford A., Aramayo R.,
RA      Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA      Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.W.;
RT      "The genome sequence of the filamentous fungus Neurospora crassa.";
RL      Nature 422:859-868(2003).
CC      -!- CAUTION: The sequence shown here is derived from an
CC      EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC      preliminary data.
CC      [1]
CC      Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC      Distributed under the Creative Commons Attribution-NoDerivs License
CC      EMBL; AABX01000368; EAA30363.1; -; Genomic DNA.
DR      GO; GO:0008080; F:N-acetyltransferase activity; IEA.
DR      InterPro; IPR000182; GCN5acetyl_trans.
DR      Pfam; PF00583; Acetyltransf_1; 1.
KW      Hypothetical protein.
SQ      SEQUENCE      213 AA;      24470 MW;      75FF1BF4E5AA4B39 CRC64;

Query Match      53.5%;      Score 46;      DB 2;      Length 213;
Best Local Similarity      66.7%;      Pred. No. 46;
Matches      8;      Conservative      1;      Mismatches      3;      Indels      0;      Gaps      0;

QY      3 HHYLYNGRTATR 14
|      ||||| :|
Db      176 HRYLYNGNSAYR 187

RESULT 28
Q5B387 EMENI
ID      Q5B387 EMENI      PRELIMINARY;      PRT;      213 AA.
AC      Q5B387;
DT      26-APR-2005, integrated into UniProtKB/TrEMBL.
DT      26-APR-2005, sequence version 1.
DT      07-MAR-2006, entry version 6.
DE      Hypothetical protein.
GN      ORFNames=AN4993.2;
OS      Aspergillus nidulans FGSC A4.
OC      Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC      Eurotiales; Trichocomaceae; Emericella.
OX      NCBI_TaxID=227321;
RN      [1]
RP      NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC      STRAIN=FGSC 4;
RX      PubMed=16372000; DOI=10.1038/nature04341;
RA      Galagan J.E., Calvo S.E., Cuomo C., Ma L.-J., Wortman J.R.,
RA      Batzoglou S., Lee S.-I., Bastuerkmen M., Spevak C.C., Clutterbuck J.,
RA      Kapitonov V., Jurka J., Scaccocchio C., Farman M., Butler J.,
RA      Purcell S., Harris S., Braus G.H., Draht O., Busch S., D'Enfert C.,
RA      Bouchier C., Goldman G.H., Bell-Pedersen D., Griffiths-Jones S.,
RA      Doonan J.H., Yu J., Vienken K., Pain A., Freitag M., Selker E.U.,
RA      Archer D.B., Penalva M.A., Oakley B.R., Momany M., Tanaka T.,
RA      Kumagai T., Asai K., Machida M., Nierman W.C., Denning D.W.,
RA      Caddick M., Hynes M., Paolletti M., Fischer R., Miller B.L., Dyer P.S.,
RA      Sachs M.S., Osmani S.A., Birren B.W.;
RT      "Sequencing of Aspergillus nidulans and comparative analysis with A.
RT      fumigatus and A. oryzae.";
RL      Nature 438:1105-1115(2005).
CC      Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC      Distributed under the Creative Commons Attribution-NoDerivs License
CC      EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC      preliminary data.

```



CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
-----  
CC EMBL; AACD0100084; EAA61071.1; -: Genomic DNA.  
DR GO; GO:0008080; F:N-acetyltransferase activity; IEA.  
DR InterPro; IPR000182; GCN5acetyl\_trans.  
DR Pfam; PF00593; Acetyltransf\_1; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 213 AA; 24228 MW; 8BDB1FBADD83BCBD CRC64;  
  
Query Match 53.5%; Score 46; DB 2; Length 213;  
Best Local Similarity 66.7%; Pred. No. 46;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
QY 3 HHVYLNGRTATR 14  
| | | | | : | |  
Db 156 HRYLNGNSAYR 167  
  
RESULT 29  
Q4WXL6 ASPFU  
ID Q4WXL6 ASPFU PRELIMINARY; PRT; 279 AA.  
AC Q4WXL6;  
DT 05-JUL-2005, integrated into UniProtKB/TrEMBL.  
DT 05-JUL-2005, sequence version 1.  
DT 07-MAR-2006, entry version 6.  
DE Acetyltransferase, GNAT family, putative.  
OS ORFNames=AFu3g09940;  
GN Aspergillus fumigatus (Sartorya fumigata).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5085;  
RN [1]  
NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RP STRAIN=AF293 / CBS 101355 / FGSC A1100;  
RX PubMed=16372009; DOI=10.1038/nature04332;  
RA Nierman W.C., Pain A., Anderson M.J., Wortman J.R., Kim H.S.,  
RA Arroyo J., Berriman M., Abe K., Archer D.B., Bernejo C., Bennett J.W.,  
RA Bowyer P., Chen D., Collins M., Coulson R., Davies R., Dyer P.S.,  
RA Farman M., Fedorova N., Fedorova N.D., Feldblyum T.V., Fischer R.,  
RA Foster N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,  
RA Goldman G.H., Gomi K., Griffith-Jones S., Gwilliam R., Haas B.J.,  
RA Haas H., Harris D.E., Horiuchi H., Huang J., Humphray S., Jimenez J.,  
RA Keller N., Khouri H., Kitamoto K., Kobayashi T., Konzack S.,  
RA Kulkarni R., Kumagai T., Lafont A., Latge J.-P., Li W., Lord A.,  
RA Lu C., Majeros W.H., May G.S., Miller B.L., Mohamoud Y., Molina M.,  
RA Monod M., Mouyna I., Mulligan S., Murphy L.D., O'Neil S., Paulsen I.,  
RA Penalva M.A., Perlea M., Price C., Pritchard B.L., Quail M.A.,  
RA Rabinowitsch E., Rawlins N., Rajandream M.A., Reichard U.,  
RA Renauld H., Robson G.D., Rodriguez de Cordoba S., Rodriguez-Pena J.M.,  
RA Ronning C.M., Rutter S., Salzberg S.L., Sanchez M.,  
RA Sanchez-Ferrero J.C., Saunders D., Seeger K., Squares R., Squares S.,  
RA Takeuchi M., Tekala F., Turner G., Vazquez de Aldana C.R., Weidman J.,  
RA White O., Woodward J.R., Yu J.-H., Fraser C.M., Galagan J.E., Asai K.,  
RA Machida M., Hall N., Barrell B.G., Denning D.W.;  
RA "Genomic sequence of the pathogenic and allergenic filamentous fungus  
RT Aspergillus fumigatus";  
RL Nature 438:1151-1156(2005).  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
-----  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
-----  
CC EMBL; AAH0100002; EAL92587.1; -: Genomic DNA.  
DR GO; GO:0008080; F:N-acetyltransferase activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
KW Complete proteome; Transferase.  
SQ SEQUENCE 279 AA; 31318 MW; 193FB4FD8EA2407C CRC64;  
  
Query Match 53.5%; Score 46; DB 2; Length 279;

Best Local Similarity 66.7%; Pred. No. 60;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
QY 3 HHVYLNGRTATR 14  
| | | | | : | |  
Db 231 HRYLNGNSAYR 242  
  
RESULT 30  
Q5DZ88 VIBF1  
ID Q5DZ88 VIBF1 PRELIMINARY; PRT; 301 AA.  
AC Q5DZ88;  
DT 15-MAR-2005, integrated into UniProtKB/TrEMBL.  
DT 15-MAR-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Hypothetical protein.  
OS OrderedLocusNames=VFA0838;  
GN Vibrio fischeri (Strain ATCC 700601 / ES114).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=312309;  
RN [1]  
NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RP PubMed=15703294; DOI=10.1073/pnas.0409900102;  
RX Ruby E.G., Urbanowski M., Campbell J., Dunn A., Faini M., Gunsalus R.,  
RA Lostroh P., Lupp C., McCann J., Millikan D., Schaefer A., Stabb E.,  
RA Stevens A., Visick K., Whistler C., Greenberg E.P.;  
RT "Complete genome sequence of Vibrio fischeri: a symbiotic bacterium  
RT with pathogenic congeners";  
RL Proc. Natl. Acad. Sci. U.S.A. 102:3004-3009(2005).  
CC -----  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
-----  
CC EMBL; CP000021; AA087908.1; -: Genomic DNA.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 301 AA; 34109 MW; D85E45C8469A9693 CRC64;  
  
Query Match 53.5%; Score 46; DB 2; Length 301;  
Best Local Similarity 61.5%; Pred. No. 65;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 TWHYTLNGRTAT 13  
| : | | | | : | | | |  
Db 55 TPXHYTFNGKAT 67  
  
RESULT 31  
Q8BSA0 MOUSE  
ID Q8BSA0 MOUSE PRELIMINARY; PRT; 349 AA.  
AC Q8BSA0;  
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
DT 01-MAR-2003, sequence version 1.  
DT 07-FEB-2006, entry version 20.  
DE 12 days embryo embryonic body between diaphragm region and neck cDNA,  
DE RIKEN full-length enriched library, clone:9430050G04  
DE product:sialyltransferase 5, full insert sequence.  
OS Name=St3gal2; Synonyms=Slat5;  
GN Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridea; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
NUCLEOTIDE SEQUENCE.  
RP TISSUE=Embryonic body between diaphragm region and neck;  
RC STRAIN=C57BL/6J;  
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6875(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning";  
RL Methods Enzymol. 303:19-44(1999).  
RN [2]  
NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RC PubMed=16141072; DOI=10.1126/science.1112014;  
 RX Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N., Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K., Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M., Davis M.J., Wilming L.G., Aidinis V., Allen J.E., Ambesi-Impombato A., Apweiler K.W., Aturaliya R.N., Bailey T.L., Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M., Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R., Crowe M.L., Dalla E., Dalrymple B.P., De Bono B., Della Gatta G., Di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G., Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M., Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E., Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N., Hill D., Huminiecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T., Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H., Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K., Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J., Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L., Matsuda H., Matsuura S., Mulder N., Nakano N., Nakaguchi H., Ng P., Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O., Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G., Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M., Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C., Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y., Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B., Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K., Tammojä K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A., Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K., Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C., Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J., Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y., Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T., Iida J., Inamori K., Itoh M., Kato T., Kawaji H., Kawagashira N., Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N., Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S., Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J., Hayashizaki Y.;  
 RL Science 309:1559-1563(2005).  
 [3]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RG RIKEN Genome Exploration Research Group, and Genome Science Group (Genome Network Core Team) and the FANTOM Consortium;  
 RG "Antisense Transcription in the Mammalian Transcriptome."; Science 309:1564-1566(2005).  
 [4]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S., Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H., Yagi K., Tomaru N., Hasegawa Y., Nogami A., Schonbach C., Gojobori T., Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S., Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L., Konegaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Nagashima T., Numata K., Okido T., Pavan W.J., Persea G., Pesole G., Petrovsky N., Pillai R., Pontius J.D., Qi D., Ramachandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M., Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,

RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N., Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K., Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."; Nature 420:563-573(2002).  
 [5]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schriml L.M., Staebli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., De Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohtsuki S., Hayashizaki Y.;  
 RL "Functional annotation of a full-length mouse cDNA collection."; Nature 409:685-690(2001).  
 [6]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes."; Genome Res. 10:1617-1630(2000).  
 [7]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and neck;  
 RX PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Embryonic body between diaphragm region and

```
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AK034863; BAC28859.1; -; mRNA.
CC Ensembl; ENSMUSG000000031749; Mus musculus.
CC MGI; MGI:99427; St3gal2.
CC GO; GO:00030173; C: integral to Golgi membrane; IEA.
CC GO; GO:0008373; F: sialyltransferase activity; IEA.
CC GO; GO:0006486; P: protein amino acid glycosylation; IEA.
CC InterPro; IPR001675; Glyco_trans_29.
CC InterPro; IPR012163; Sialyl_trans.
CC Pfam; PF00777; Glyco_transf_29; 1.
CC PIRSF; PIRSF005557; Sialyl_trans; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 349 AA; 40033 MW; 372926AE6379186E CRC64;

Query Match 53.5%; Score 46; DB 2; Length 349;
Best Local Similarity 63.6%; Pred. No. 75;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHYYLNGRTA 12
DB 306 WHYYWENRYA 316

RESULT 32
Q5ZJJO.CHICK PRELIMINARY; PRT; 349 AA.
AC Q5ZJJO;
DT 23-NOV-2004, integrated into UniProtKB/TREMBL.
DT 23-NOV-2004, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Hypothetical protein.
GN ORFNames=RCJMB04_17119;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX EMBL; AJ720444; C: integral to Golgi membrane; IEA.
DR GO; GO:0030173; C: integral to Golgi membrane; IEA.
DR GO; GO:0008373; F: sialyltransferase activity; IEA.
DR GO; GO:0016740; F: transferase activity; IEA.
DR GO; GO:0006486; P: protein amino acid glycosylation; IEA.
DR InterPro; IPR001675; Glyco_trans_29.
DR InterPro; IPR012163; Sialyl_trans.
DR Pfam; PF00777; Glyco_transf_29; 1.
DR PIRSF; PIRSF005557; Sialyl_trans; 1.
KW Transferase.
SQ SEQUENCE 349 AA; 39683 MW; B14F4CEB464F0456 CRC64;

Query Match 53.5%; Score 46; DB 2; Length 349;
Best Local Similarity 63.6%; Pred. No. 75;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHYYLNGRTA 12
DB 306 WHYYWENRYA 316

RESULT 34
SIA4B.HUMAN STANDARD; PRT; 350 AA.
AC Q16842; O00654;
DT 11-JUL-2002, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1996, sequence version 1.
DT 07-FEB-2006, entry version 49.
DE CMP-N-acetylneuraminase-beta-galactosamide-alpha-2,3-sialyltransferase
DE (EC 2.4.99.-) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha
DE 2,3-ST) (Gal-NAC6S) (Gal-beta-1,3-GalNAc-alpha-2,3-sialyltransferase)
DE (ST3GalA.2) (SIAT4-B) (ST3Gal II).
GN Name=ST3GAL2; Synonyms=SIAT4B;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RC TISSUE=Liver;
RX MEDLINE=97079181; PubMed=8920913; DOI=10.1006/bbrc.1996.1660;
```

```
QY 2 WHYYLNGRTA 12
DB 306 WHYYWENRYA 316

RESULT 33
Q70D58.CHICK PRELIMINARY; PRT; 349 AA.
AC Q70D58;
DT 05-JUL-2004, integrated into UniProtKB/TREMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Alpha2,3-sialyltransferase.
GN Name=ST3GAL-II;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15843597; DOI=10.1093/glycob/cwi063;
RA Harduin-Leperc A., Mollicone R., Delannoy P., Oriol R.;
RT "The animal sialyltransferases and sialyltransferase-related genes: a
RT phylogenetic approach.";
RL Glycobiology 15:805-817 (2005).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AJ585761; CAE51385.2; -; mRNA.
DR GO; GO:0030173; C: integral to Golgi membrane; IEA.
DR GO; GO:0008373; F: sialyltransferase activity; IEA.
DR GO; GO:0016740; F: transferase activity; IEA.
DR GO; GO:0006486; P: protein amino acid glycosylation; IEA.
DR InterPro; IPR001675; Glyco_trans_29.
DR InterPro; IPR012163; Sialyl_trans.
DR Pfam; PF00777; Glyco_transf_29; 1.
DR PIRSF; PIRSF005557; Sialyl_trans; 1.
KW Transferase.
SQ SEQUENCE 349 AA; 39683 MW; B14F4CEB464F0456 CRC64;

Query Match 53.5%; Score 46; DB 2; Length 349;
Best Local Similarity 63.6%; Pred. No. 75;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHYYLNGRTA 12
DB 306 WHYYWENRYA 316

RESULT 34
SIA4B.HUMAN STANDARD; PRT; 350 AA.
AC Q16842; O00654;
DT 11-JUL-2002, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1996, sequence version 1.
DT 07-FEB-2006, entry version 49.
DE CMP-N-acetylneuraminase-beta-galactosamide-alpha-2,3-sialyltransferase
DE (EC 2.4.99.-) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha
DE 2,3-ST) (Gal-NAC6S) (Gal-beta-1,3-GalNAc-alpha-2,3-sialyltransferase)
DE (ST3GalA.2) (SIAT4-B) (ST3Gal II).
GN Name=ST3GAL2; Synonyms=SIAT4B;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RC TISSUE=Liver;
RX MEDLINE=97079181; PubMed=8920913; DOI=10.1006/bbrc.1996.1660;
```

RA Kim Y.-J., Kim K.-S., Kim S.-H., Kim C.-H., Ko J.H., Choe I.-S.,  
RA Tsuji S., Lee Y.-C.,  
RT "Molecular cloning and expression of human Gal beta 1,3GalNAc alpha  
RL 2,3-sialyltransferase (hST3Gal II).";  
RL Biochem. Biophys. Res. Commun. 228:324-327(1996).  
[2]  
RN NUCLEOTIDE SEQUENCE [mRNA], AND TISSUE SPECIFICITY.  
RP MEDLINE=97403982; PubMed=9266697;  
RX Giordanengo V., Lefebvre J.-C.;  
RT "Cloning and expression of cDNA for a human Gal beta1-3GalNAc  
RL alpha2,3-sialyltransferase from the CEM T cell line.";  
RL Eur. J. Biochem. 247:558-566(1997).  
RN [3]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE mRNA].  
RC TISSUE=Lung;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzyzanski M.I., Skalska U., Smallus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences".  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -1- FUNCTION: It may be responsible for the synthesis of the sequence  
CC NeuAc-alpha-2,3-Gal-beta-1,3-GalNAc- found in terminal  
CC carbohydrate groups of certain glycoproteins, oligosaccharides and  
CC glycolipids. SIAT4A and SIAT4B sialylate the same acceptor  
CC substrates but exhibit different Km values.  
CC -1- CATALYTIC ACTIVITY: CMP-N-acetylneuraminatate + beta-D-galactosyl-  
CC 1,3-N-acetyl-alpha-D-galactosaminyl-R = CMP + alpha-N-  
CC acetylneuraminyl-2,3-beta-D-galactosyl-1,3-N-acetyl-alpha-D-  
CC galactosaminyl-R.  
CC -1- PATHWAY: Glycosylation.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
CC form in trans cisternae of Golgi. Soluble form in body fluids (By  
CC similarity).  
CC -1- TISSUE SPECIFICITY: Highly expressed in skeletal muscle and heart  
CC and to a much lesser extent in brain, placenta, liver and  
CC pancreas. Scarcely detectable in lung and kidney.  
CC -1- PWM: The soluble form derives from the membrane form by  
CC proteolytic processing (By similarity).  
CC -1- SIMILARITY: Belongs to the glycosyltransferase 29 family.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; U63090; AAB40389.1; -; mRNA.  
CC EMBL; X96667; CAA65447.1; -; mRNA.  
CC EMBL; BC036777; AAH36777.1; -; mRNA.  
CC FIRM; JC5251; JC5251.  
CC Ensembl; ENSG00000157350; Homo sapiens.  
CC HGNC; HGNC:10863; ST3GAL2.  
CC MIM; 607188; Gene.  
CC GO; GO:0003836; F-beta-galactoside alpha-2,3-sialyltransferase. .; TAS.  
CC GO; GO:0006040; P:amino sugar metabolism; TAS.  
CC GO; GO:0006464; P:protein modification; TAS.  
CC InterPro; IPR001675; Glyco\_trans\_29.  
CC InterPro; IPR012163; Sialyl\_trans.  
CC Pfam; PF00777; Glyco\_transf\_29; 1.  
CC PIRSF; PIRSF005557; Sialyl\_trans; 1.  
CC

KW Glycoprotein; Glycosyltransferase; Golgi stack; Membrane;  
KW Signal-anchor; Transferase; Transmembrane.  
FT CHAIN 1 350  
FT CMP-N-acetylneuraminatate-beta-  
FT galactosamide-alpha-2,3-  
FT sialyltransferase.  
FT /FTID=PRO 0000149258.  
FT Cytoplasmic (Potential).  
FT Signal-anchor for type II membrane  
FT protein (Potential).  
FT Lumenal (Potential).  
FT TOPO\_DOM 28 350  
FT CARBOHYD 92 92  
FT N-linked (GlcNAc. .) (Potential).  
FT DISULFID 152 291  
FT By similarity.  
SQ SEQUENCE 350 AA; 40173 MW; E7E40CF26D9CB725 CRC64;  
  
Query Match 53.5%; Score 46; DB 1; Length 350;  
Best Local Similarity 63.6%; Pred. No. 76;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 WHHYLNGRTA 12  
DB 307 WHHYWENRYA 317  
||||:|  
  
RESULT 35  
SIA4B PANTR STANDARD; PRT; 350 AA.  
AC Q6KB58;  
DT 15-FEB-2005, integrated into UniProtKB/Swiss-Prot.  
DT 05-JUL-2004, sequence version 1.  
DT 07-FEB-2006, entry version 15.  
DE CMP-N-acetylneuraminatate-beta-galactosamide-alpha-2,3-sialyltransferase  
DE (EC 2.4.99.-) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha  
DE 2,3-ST) (Gal-beta-1,3-GalNAc-alpha-2,3-sialyltransferase)  
DE (ST3GAL2) (SIAT4-B) (ST3Gal II).  
GN Name=ST3GAL2; Synonyms=SIAT4B;  
OS Pan troglodytes (Chimpanzee).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Pan.  
OX NCBI\_TaxID=9598;  
RN [1]  
RA Harduin-Leperc A., Martinez-Duncker I., Mollicone R., Delannoy P.,  
RA Ortol R.;  
RT "Phylogeny of sialyltransferases";  
RL Submitted (JUN-2004) to the EMBL/GenBank/DBSJ databases.  
CC -1- FUNCTION: It may be responsible for the synthesis of the sequence  
CC NeuAc-alpha-2,3-Gal-beta-1,3-GalNAc- found in terminal  
CC carbohydrate groups of certain glycoproteins, oligosaccharides and  
CC glycolipids. SIAT4A and SIAT4B sialylate the same acceptor  
CC substrates but exhibit different Km values (By similarity).  
CC -1- CATALYTIC ACTIVITY: CMP-N-acetylneuraminatate + beta-D-galactosyl-  
CC 1,3-N-acetyl-alpha-D-galactosaminyl-R = CMP + alpha-N-  
CC acetylneuraminyl-2,3-beta-D-galactosyl-1,3-N-acetyl-alpha-D-  
CC galactosaminyl-R.  
CC -1- PATHWAY: Glycosylation.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound  
CC form in trans cisternae of Golgi. Soluble form in body fluids (By  
CC similarity).  
CC -1- PWM: The soluble form derives from the membrane form by  
CC proteolytic processing (By similarity).  
CC -1- SIMILARITY: Belongs to the glycosyltransferase 29 family.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; AJ744804; CAG32840.1; -; mRNA.  
CC InterPro; IPR001675; Glyco\_trans\_29.  
CC InterPro; IPR012163; Sialyl\_trans.  
CC Pfam; PF00777; Glyco\_transf\_29; 1.  
CC PIRSF; PIRSF005557; Sialyl\_trans; 1.  
KW Glycoprotein; Glycosyltransferase; Golgi stack; Membrane;

KW	Signal-anchor; Transferase; Transmembrane.
FT	CHAIN 1 350
FT	CMP-N-acetylneuraminase-beta-
FT	galactosamide-alpha-2,3-
FT	sialyltransferase.
FT	/FTid=PRO.0000149260.
FT	Cytoplasmic (Potential).
FT	TOPO_DOM 1 6
FT	TRANSMEM 7 27
FT	Signal-anchor for type II membrane
FT	protein (Potential).
FT	TOPO_DOM 28 350
FT	CARBHYD 211 211
FT	Lumenal (Potential).
FT	DISULFID 152 291
FT	N-linked (GlcNAc. . .) (Potential).
FT	By similarity.
SQ	SEQUENCE 350 AA; 40074 MW; FCE9932A2D9CB73A CRC64;
Query Match	53.5%; Score 46; DB 1; Length 350;
Best Local Similarity	63.6%; Pred. No. 76;
Matches	7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY	2 WHHYLNGRTA 12
	:
Db	307 WHHYENNRYA 317
RESULT 36	
SI44B_RAT	
ID	SI44B_RAT STANDARD; PRT; 350 AA.
AC	Q11205;
DT	01-OCT-1996, integrated into UniProtKB/Swiss-Prot.
DT	01-OCT-1996, sequence version 1.
DT	07-FEB-2006, entry version 40.
DE	CMP-N-acetylneuraminase-beta-galactosamide-alpha-2,3-sialyltransferase
DE	(EC 2.4.99.-) (Beta-galactoside alpha-2,3-sialyltransferase) (Alpha
DE	2,3-ST) (Gal-NAC6S) (Gal-beta-1,3-GalNAc-alpha-2,3-sialyltransferase)
DE	(ST3GALA.2) (SIAT4-B) (ST3Gal II).
GN	Names:Stgal2; Synonyms:SIAT4b, Siat5;
GN	Rattus norvegicus (Rat).
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC	Muroidea; Muridae; Murinae; Rattus.
OC	NCBI_TaxID=10116;
RN	[1]
RP	NUCLEOTIDE SEQUENCE [MRNA].
RP	TISSUE=Brain;
RX	MEDLINE=94193584; PubMed=8144500;
RA	Lee Y.-C., Kojima N., Wada E., Kurosawa N., Nakaoka T., Hamamoto T.,
RA	Tsuiji S.;
RT	"Cloning and expression of cDNA for a new type of Gal beta 1,3GalNAc
RT	alpha 2,3-sialyltransferase.";
RL	J. Biol. Chem. 269:10028-10033(1994).
CC	-1- FUNCTION: It may be responsible for the synthesis of the sequence
CC	NeuAc-alpha-2,3-Gal-beta-1,3-GalNAc- found in terminal
CC	carbohydrate groups of certain glycoproteins, oligosaccharides and
CC	glycolipids. SIAT4A and SIAT4B sialylate the same acceptor
CC	substrates but exhibit different Km values.
CC	-1- CATALYTIC ACTIVITY: CMP-N-acetylneuraminase + beta-D-galactosyl-
CC	1,3-N-acetyl-alpha-D-beta-D-galactosaminyl-R = CMP + alpha-N-
CC	acetylneuraminyl-2,3-beta-D-galactosyl-1,3-N-acetyl-alpha-D-
CC	galactosaminyl-R.
CC	-1- PATHWAY: Glycosylation.
CC	-1- SUBCELLULAR LOCATION: Type II membrane protein. Membrane-bound
CC	form in trans cisternae of Golgi. Soluble form in body fluids.
CC	-1- PTM: The soluble form derives from the membrane form by
CC	proteolytic processing.
CC	-1- SIMILARITY: Belongs to the glycosyltransferase 29 family.
CC	Copyrighted by the UniProt Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a>
CC	Distributed under the Creative Commons Attribution-NoDerivs License
CC	-----
CC	EMBL: X76988; CAA54293.1; -; mRNA.
CC	PIR: B54420; B54420.
DR	Ensembl: ENSRNOG0000017932; Rattus norvegicus.
DR	RGD: 68413; Siat4b.
DR	InterPro: IPR0161675; Glyco_trans_29.
DR	InterPro: IPR012163; Sialyl trans.

Query Match 53.5%; Score 46; DB 2; Length 350;  
 Best Local Similarity 63.6%; Pred. No. 76;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WHYYLNGRTA 12  
 |||||  
 DB 307 WHYWNRYA 317

RESULT 38  
 Q8BPLO MOUSE  
 ID QBEPLO MOUSE PRELIMINARY; PRT; 350 AA.  
 AC QBEPLO;  
 DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
 DT 01-MAR-2003, sequence version 1.  
 DT 07-FEB-2006, entry version 22.  
 DE O day neonate eyeball cDNA, RIKEN full-length enriched library,  
 DE clone neonate eyeball cDNA, RIKEN full-length enriched library,  
 DE DE E130314A03 product: sialyltransferase 5, full insert sequence  
 DE (ST3 beta-galactoside alpha-2,3-sialyltransferase 2) (Adult male  
 DE cerebellum cDNA, RIKEN full-length enriched library, clone:1520403I17  
 DE product:sialyltransferase 5, full insert sequence).  
 GN Name:St3gal2; Synonyms:Siat5;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muroidae; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RX NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Cerebellum, and Eyeball;  
 RC MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RX PubMed=16141072; DOI=10.1038/science.112014;  
 RA Carninci P., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
 RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Hill D., Humnietek L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
 RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 RA Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,  
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
 RA Petrovsky N., Piazza S., Reed J.C., Reid J.F., Ring B.Z., Ringwald M.,  
 RA Rost B., Ruan Y., Salzberg S.L., Sadelin A., Schneider C.,  
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
 RA Shibata Y., Shimada K., Shimada K., Silva D., Sinclair B.,  
 RA Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,  
 RA Tamjoo K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawaehima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,

RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RT "The transcriptonal landscape of the mammalian genome.";  
 RL Science 309:1559-1563(2005).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Eyeball;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RG. RIKEN Genome Exploration Research Group, and Genome Science Group  
 RG (Genome Network Core Team) and the FANTOM Consortium;  
 RT "Antisense Transcription in the Mammalian Transcriptome";  
 RL Science 309:1564-1566(2005).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Eyeball;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuta M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sadelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sakai D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs";  
 RL Nature 420:563-573(2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Eyeball;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Flaischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Bozell D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Eyeball;







```

RESULT 41
Q4AXU3_9BURK
ID Q4AXU3_9BURK PRELIMINARY; PRT; 892 AA.
AC Q4AXU3_
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 2.
DE Transposase Tn3.
GN ORFNames=BprodRAFT 4286;
OS Polaromonas sp. JS666.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Comamonadaceae; Polaromonas.
OC NCBI_TaxID=296591;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=JS666;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Polaromonas sp.
RT JS666.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=JS666;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft assembly of Polaromonas sp. JS666.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=JS666;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AAFQ2000007; EAM38976.1; -; Genomic_DNA.
CC GO; GO:0003677; F:DNA binding; IEA.
CC GO; GO:0004803; F:transposase activity; IEA.
CC GO; GO:0006313; P:DNA transposition; IEA.
CC SEQUENCE 892 AA; 101919 MW; 14E1EFD8A791B848 CRC64;

Query Match 53.5%; Score 46; DB 2; Length 892;
Best Local Similarity 61.5%; Pred. No. 2e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 WHHYLNGRTATR 14
| | | | |
Db 863 WRHVLNGRYA 875

RESULT 42
O80438_ARATH
ID O80438_ARATH PRELIMINARY; PRT; 190 AA.
AC O80438_
DT 01-NOV-1998, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1998, sequence version 1.
DT 07-FEB-2006, entry version 23.
DE Putative acetyltransferase (At2g38130).
GN OrderedLocusNames=At2g38130;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eucosids II; Brassicales; Brassicaceae; Arabidopsi.

```

```
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Rounsley S.D., Kaul S., Lin X., Ketchum K.A., Crosby M.L.,
RA Brandon R.C., Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D.,
RA Somerville C.R., Venter J.C.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Town C.D., Kaul S.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Ecker J.R.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Ecker J.R.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AC003028; AAC27162.1; -; Genomic_DNA.
DR EMBL; BT010841; AAR24208.1; -; mRNA.
DR EMBL; BT012615; AAT06434.1; -; mRNA.
DR PIR; T01245; T01245.
DR TAIR; At2g38130; -.
DR GO; GO:0008080; F.N-acetyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000182; GCN5acetyl trans.
DR Pfam; PF00593; Acetyltransf_1; 1.
KW Transferase.
SQ SEQUENCE 190 AA; 22123 MW; 124BD9D4D0A491 CRC64;

Query Match 52.3%; Score 45; DB 2; Length 190;
Best Local Similarity 66.7%; Pred. No. 59;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 HHVYLNGRATR 14
DB 147 YHLYNGMDAFR 158

RESULT 43
Q2URZ0 ASPOR PRELIMINARY; PRT; 448 AA.
AC Q2URZ0;
DT 24-JAN-2006, integrated into UniprotKB/TrEMBL.
DT 24-JAN-2006, sequence version 1.
DT 07-MAR-2006, entry version 3.
DE Predicted protein.
GN ORFNames=AO090005000633;
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=RIB 40;
RC PubMed=16372010; DOI=10.1038/nature04300;
RA Machida M., Asai K., Sano M., Tanaka T., Kumagai T., Terai G.,
RA Kusumoto K., Arima T., Akita O., Kashiwagi Y., Abe K., Gomi K.,
RA Horiuchi H., Kitamoto K., Kobayashi T., Takeuchi M., Denning D.W.,
RA Galagan J.E., Nierman W.C., Yu J., Archer D.B., Bennett J.W.,
RA Bhatnagar D., Cleveland T.E., Fedorova N.D., Gotoh O., Horikawa H.,
RA Hosoyama A., Ichinomiya M., Igazashi R., Iwashita K., Juvvadi P.R.,
RA Kato M., Kato Y., Kin T., Kokubun A., Maeda H., Maeyama N.,
RA Maruyama J., Nagaoka H., Nakajima T., Oda K., Okada K., Paulsen I.,
RA Sakamoto K., Sawano T., Takahashi M., Takase K., Terabayashi Y.,
RA Wortman J.R., Yamada O., Yamagata Y., Anazawa H., Hata Y., Koide Y.,
RA Komori T., Koyama Y., Minetoki T., Suharnan S., Tanaka A., Isono K.,
RA Kuhara S., Ogasawara N., Kikuchi H.;
```

```
RT "Genome sequencing and analysis of Aspergillus oryzae.";
RL Nature 438:1157-1161(2005).
CC -!- SUBCELLULAR LOCATION: Membrane; multi-pass membrane protein (By
CC similarity).
CC -----
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AP007151; BAB55675.1; -; Genomic_DNA.
DR EMBL; Transmembrane; Transport.
SQ SEQUENCE 448 AA; 49144 MW; 740F5AB85B3C9790 CRC64;

Query Match 52.3%; Score 45; DB 2; Length 448;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 WHHYL 7
DB 184 WHHYL 189

RESULT 44
Q33X5 SPAJD PRELIMINARY; PRT; 558 AA.
AC Q33X5;
DT 06-DEC-2005, integrated into UniprotKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Names=hpa;
OS Spalax judaei (Blind subterranean mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=134510;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT cloning and identification of a novel splice variant.";
RL Cloning. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT cloning and identification of a novel splice variant.";
RL Cloning. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC -----
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085494; CAJ30021.1; -; mRNA.
SQ SEQUENCE 558 AA; 62737 MW; 07BAF8F5849EE7 CRC64;

Query Match 52.3%; Score 45; DB 2; Length 558;
Best Local Similarity 80.0%; Pred. No. 1.8e+02;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 YLNGRTATR 14
DB 313 YLNGRIATK 322

RESULT 45
Q4RHG9 TETNG PRELIMINARY; PRT; 612 AA.
AC Q4RHG9;
DT 19-JUL-2005, integrated into UniprotKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE Chromosome 3 SCAF15050, whole genome shotgun sequence. (Fragment).
```

GN ORFNames=GSTENG0034356001;  
OS Tetraodon nigroviridis (Green puffer).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
OC Tetraontoidea; Tetraodontidae; Tetraodon.  
OX NCBI\_TaxID=99883;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15496914; DOI=10.1038/nature03025;  
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,  
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,  
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,  
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,  
RA Blemont C., Skalli Z., Cottolico L., Poullain J., De Berardinis V.,  
RA Craud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,  
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,  
RA Kallis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,  
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
RA Wancker P., Lander E.S., Weissbach J., Roest Crollius H.;  
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals  
RT the early vertebrate proto-karyotype";  
RL Nature 431:946-957(2004).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RG Genoscope; Whitehead Institute Centre for Genome Research;  
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -!- FAMILY: Belongs to the ATP-dependent AMP-binding enzyme  
CC family.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR ENBL; CAE01015050; CAG12163.1; -; Genomic DNA.  
DR GO; GO:0003824; F: catalytic activity; IEA.  
DR GO; GO:0008152; P: metabolism; IEA.  
DR InterPro; IPR000873; AMP-bind.  
DR Pfam; PF00501; AMP-binding; 1.  
DR PRINTS; PR00154; AMPBINDING.  
DR PROSITE; PS00455; AMP BINDING; 1.  
FT NON\_TER 1 1  
FT NON\_TER 612 612  
SQ SEQUENCE 612 AA; 67282 MW; BC8AGE6727D0E297 CRC64;  
Query Match 52.3%; Score 45; DB 2; Length 612;  
Best Local Similarity 50.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 1 TWHHYUNGRTATR 14  
Db : : : : :  
9 TWKYFYIAARTAKR 22  
RESULT 46  
Q6PY34\_9POTY PRELIMINARY; PRT; 3050 AA.  
AC Q6PY34;  
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.  
DT 05-JUL-2004, sequence version 1.  
DT 07-FEB-2006, entry version 9.  
DE Polypeptide.  
OS Hordeum mosaic virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;  
OC Rymovirus.  
OX NCBI\_TaxID=41764;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=ATCC PV81;

RX PubMed=15193921; DOI=10.1016/j.virol.2004.03.014;  
RA Stenger D.C., French R.;  
RT "Functional replacement of Wheat streak mosaic virus HC-Pro with the  
RT corresponding citron from a diverse array of viruses in the family  
RT Potyviridae";  
RL Virology 323:257-267(2004).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=ATCC PV81;  
RX PubMed=15449142; DOI=10.1007/s00705-004-0396-6;  
RA French R., Stenger D.C.;  
RT "Genome sequences of Agropyron mosaic virus and Hordeum mosaic virus  
RT support reciprocal monophyly of the genera Potyvirus and Rymovirus in  
RT the family Potyviridae";  
RL Arch. Virol. 150:299-312(2005).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR ENBL; AY623627; AAS65455.2; -; Genomic RNA.  
DR GO; GO:0019028; C: viral capsid; IEA.  
DR GO; GO:0005524; F: ATP binding; IEA.  
DR GO; GO:0008026; F: ATP-dependent helicase activity; IEA.  
DR GO; GO:0004137; F: cysteine-type endopeptidase activity; IEA.  
DR GO; GO:0003723; F: RNA binding; IEA.  
DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:0006508; P: proteolysis; IEA.  
DR GO; GO:0019079; P: viral genome replication; IEA.  
DR InterPro; IPR001410; DEAD/DEAH\_N.  
DR InterPro; IPR011545; DEAD/DEAH\_N.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR002345; Lipocalin.  
DR InterPro; IPR002540; Pept\_S30\_PotY\_P1.  
DR InterPro; IPR001730; Peptidase\_C4.  
DR InterPro; IPR001456; Peptidase\_C6.  
DR InterPro; IPR001592; Poty\_coat.  
DR InterPro; IPR001205; RNA\_pol\_P3D.  
DR InterPro; IPR007095; RNA\_pol\_P3D.  
DR InterPro; IPR007094; RNA\_pol\_Psvir.  
DR Pfam; PF00270; DEAD\_1.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF00863; Peptidase\_C4; 1.  
DR Pfam; PF00851; Peptidase\_C6; 1.  
DR Pfam; PF01577; Peptidase\_S30; 1.  
DR Pfam; PF00767; Poty\_coat; 1.  
DR Pfam; PF00680; RGRP\_1; 1.  
DR PRINTS; PR00966; NIAPOTYPTASE.  
DR SMART; SM00487; DEXdc; 1.  
DR SMART; SM00490; HELICC; 1.  
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN\_1.  
KW Polyprotein.  
SQ SEQUENCE 3050 AA; 344716 MW; 455C2C163256F888 CRC64;  
Query Match 51.7%; Score 44.5; DB 2; Length 3050;  
Best Local Similarity 69.2%; Pred. No. 1.2e+03;  
Matches 9; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
QY 2 WHHYLYN-GRAT 13  
Db : : : : :  
41 WHAYYLNAGRELT 53  
RESULT 47  
Q7V6T4\_PROMM  
ID Q7V6T4\_PROMM PRELIMINARY; PRT; 161 AA.  
AC Q7V6T4;  
DT 01-OCT-2003, integrated into UniProtKB/TrEMBL.  
DT 01-OCT-2003, sequence version 1.  
DT 07-FEB-2006, entry version 11.  
DE Putative ribosomal-protein-alanine acetyltransferase (EC 2.3.1.128).  
GN OrderedLocusNames=PMT1060;  
OS Prochlorococcus marinus (strain MIT 9313).

OC Bacteria; Cyanobacteria; Prochlorales; Prochlorococcaceae;  
OC Prochlorococcus.  
OX NCBI\_TaxID=74547;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RX MEDLINE=22825698; PubMed=12917642; DOI=10.1038/nature01947;  
RA Rocap G., Larimer F.W., Lamerdin J.E., Malfatti S., Chain P.,  
RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,  
RA Johnson Z.I., Land M.L., Lindell D., Post A.F., Regala W., Shah M.,  
RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,  
RA Webb E.A., Zinser E.R., Chisholm S.W.;  
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic  
RT niche differentiation."  
RL Nature 424:1042-1047(2003).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; BX572098; CAE21235.1; -; Genomic\_DNA.  
DR BioCyc; PMAR74547; PMT1060-MONOMER; -;  
DR GO; GO:0008415; F:acetyltransferase activity; IEA.  
DR GO; GO:0008080; F:N-acetyltransferase activity; IEA.  
DR GO; GO:0008999; F:ribosomal-protein-alanine N-acetyltransferase. .; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR InterPro; IPR000182; GCN5acetyl trans.  
DR Pfam; PF00583; Acetyltransf 1; 1.  
KW Acyltransferase; Complete proteome; Transferase.  
SQ SEQUENCE 161 AA; 17972 MW; D18ED45E9E0C4493 CRC64;  
  
Query Match 51.2%; Score 44; DB 2; Length 161;  
Best Local Similarity 70.0%; Pred. No. 71;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
QY 3 HHYLLNGRTA 12  
||| |||  
DB 135 HHYNSGQDA 144  
  
RESULT 48  
QASB93\_TETNG  
ID Q4SB93\_TETNG PRELIMINARY; PRT; 234 AA.  
AC Q4SB93;  
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
DT 19-JUL-2005, sequence version 1.  
DT 07-FEB-2006, entry version 5.  
DE Chromosome undetermined SCAF14676, whole genome shotgun sequence.  
DE (Fragment).  
GN ORFNames=GSTENG0021075001;  
OS Tetraodon nigroviridis (Green puffer).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
OC Tetraodontidae; Tetraodontidae; Tetraodon.  
OX NCBI\_TaxID=99883;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15496914; DOI=10.1038/nature03025;  
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,  
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,  
RA Nicaud S., Jaffe D., Fisher S., Lufalla G., Dossat C., Segurens B.,  
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,  
RA Biemont C., Skalli Z., Cattolico L., Poulin J., De Berardinis V.,  
RA Craud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,  
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,  
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,  
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
RA Laudet V., Schacher V., Quetier F., Saurin W., Scarpelli C.,  
RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;  
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals  
RT the early vertebrate proto-karyotype."  
RL Nature 431:946-957(2004).  
RN [2]

RP NUCLEOTIDE SEQUENCE.  
RG Genoscope; Whitehead Institute Centre for Genome Research;  
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; CAAB01014676; CAG02089.1; -; Genomic\_DNA.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003700; F:transcription factor activity; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR001356; Homeobox.  
DR InterPro; IPR000047; HTH\_lambrepresr.  
DR Pfam; PF00046; Homeobox; 1.  
DR PRINTS; PR00024; HOMEBOX.  
DR PRINTS; PR00031; HTHREPRESSR.  
DR ProDom; PD000010; Homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX 1; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
KW DNA-binding; Homeobox; Nuclear protein.  
FT NON TER 1  
SQ SEQUENCE 234 AA; 26319 MW; E0B2315DC2EC02BB CRC64;  
  
Query Match 51.2%; Score 44; DB 2; Length 234;  
Best Local Similarity 77.8%; Pred. No. 1e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 TWHYYLNG 9  
|:|:| |  
DB 50 TYHHYNLNG 58  
  
RESULT 49  
Q404U7\_9RHOB  
ID Q404U7\_9RHOB PRELIMINARY; PRT; 236 AA.  
AC Q404U7;  
DT 27-SEP-2005, integrated into UniProtKB/TrEMBL.  
DT 27-SEP-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Peptidase M50.  
DE ORFNames=JannDRAFT\_4278;  
GN Jannaschia sp. CCS1.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;  
OC Rhodobacteraceae; Jannaschia.  
OX NCBI\_TaxID=290400;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=CCS1;  
RG US DOE Joint Genome Institute (JGI-PGF);  
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,  
RA Hammon N., Israni S., Pitluck S., Richardson P.;  
RT "Sequencing of the draft genome and assembly of Jannaschia sp. CCS1."  
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=CCS1;  
RG US DOE Joint Genome Institute (JGI-ORNL);  
RA Larimer F., Land M.;  
RT "Annotation of the draft genome assembly of Jannaschia sp. CCS1."  
CC Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AAIG01000018; EAM65398.1; -; Genomic\_DNA.

DR GO: GO:0004222; F:metalloendopeptidase activity; IEA.  
DR GO: GO:0006508; P:proteolysis; IEA.  
DR InterPro: IPR008915; Peptidase\_M50.  
DR Pfam: PF02163; Peptidase\_M50; I.  
SQ SEQUENCE 236 AA; 25641 MW; 348E8A2AD1B64F72 CRC64;

Query Match 51.2%; Score 44; DB 2; Length 236;  
Best Local Similarity 50.0%; Pred. No. 1.1e+02;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 TWHYLLNGRTAT 14  
:|:|:|:|:|  
Db 177 SWLHRFVNGATNR 190

## RESULT 50

Q5W6T7 ORYSA  
AC Q5W6T7 ORYSA PRELIMINARY; PRT; 260 AA.  
DT 07-DEC-2004, integrated into UniProtKB/TrEMBL.  
DT 07-DEC-2004, sequence version 1.  
DT 07-FEB-2006, entry version 6.  
DE Hypothetical protein P0018A03.12.  
GN Name=P0018A03.12;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Chow T.-Y., Hsing Y.-I.C., Chen C.-S., Chen H.-H., Liu S.-M.,  
RA Chao Y.-T., Chang S.-J., Chen H.-C., Chen S.-K., Chen T.-R.,  
RA Chen Y.-L., Cheng C.-H., Chung C.-I., Han S.-Y., Hsiao S.-H.,  
RA Hsiung J.-N., Hsu C.-H., Huang J.-J., Kau P.-I., Lee M.-C., Leu H.-L.,  
RA Li Y.-F., Lin S.-J., Lin Y.-C., Wu S.-W., Yu C.-Y., Yu S.-W.,  
RA Wu H.-P., Shaw J.-F., Yu Y., Rambo T., Currie J., Collura K.,  
RA Soderlund C., Wing R.;  
RA "Oryza sativa PAC P0018A03 genomic sequence."  
RL Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AC134346; AAV44183.1; -; Genomic DNA.  
DR Gramene; Q5W6T7; -;  
DR InterPro: IPR006458; DUF623\_pln.  
DR Pfam: PF04844; DUF623; 1.  
DR TIGRFAMs: TIGR01569; A\_thal\_3678; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 260 AA; 28936 MW; A23DB19E7D40C568 CRC64;

Query Match 51.2%; Score 44; DB 2; Length 260;  
Best Local Similarity 63.6%; Pred. No. 1.2e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 HHYLLNGRTAT 13  
|||||:|  
Db 56 HHYLLHGHTT 66

Search completed: June 5, 2006, 12:53:58  
Job time : 159.74 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:31:47 ; Search time 104.384 Seconds

(without alignments)  
65.702 Million cell updates/sec

Title: US-10-645-659A-9

Perfect score: 82

Sequence: 1 RPGKKVWLGETSSAY 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2599679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2599679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : A\_Geneseq\_8.\*

1: geneseqp1980s.\*

2: geneseqp1990s.\*

3: geneseqp2000s.\*

4: geneseqp2001s.\*

5: geneseqp2002s.\*

6: geneseqp2003as.\*

7: geneseqp2003bs.\*

8: geneseqp2004s.\*

9: geneseqp2005s.\*

10: geneseqp2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	15	ADL16423	Adl16423 Human hep
2	82	100.0	15	ADR88215	Adr88215 Human hep
3	82	100.0	15	ADT78182	Adt78182 Functiona
4	82	100.0	15	ADU71218	Adu71218 Human hep
5	82	100.0	15	AEA42431	Aea42431 Human hep
6	82	100.0	386	ADR88207	Adr88207 Human mat
7	82	100.0	386	ADT78174	Adt78174 45kDa sub
8	82	100.0	386	ADY27057	Ady27057 Heparanas
9	82	100.0	386	ADZ18995	Adz18995 Human hep
10	82	100.0	386	AEA42423	Aea42423 Human mat
11	82	100.0	460	ADY27061	Ady27061 Heparanas
12	82	100.0	486	AE887589	Ae887589 Human hep
13	82	100.0	492	ADZ18996	Adz18996 Hep106 co
14	82	100.0	493	AE887562	Ae887562 Human hep
15	82	100.0	495	ADZ18999	Adz18999 Hep109 co
16	82	100.0	497	AE887587	Ae887587 Human hep
17	82	100.0	501	ADZ19000	Adz19000 HepGS3 co
18	82	100.0	507	ADZ19005	Adz19005 HepGS6 co
19	82	100.0	508	ADY27058	Ady27058 Human ina
20	82	100.0	526	ADZ19006	Adz19006 HepHyalur
21	82	100.0	527	AB807815	Ab807815 Chicken s
22	82	100.0	527	ABW02018	Abw02018 Chimeric
23	82	100.0	527	ADO63825	Ado63825 Chimeric

24	82	100.0	527	9	ADZ19004	Adz19004 HepGS4 co
25	82	100.0	530	2	AAV34173	Aav34173 Human pre
26	82	100.0	532	2	AAV17083	Aay17083 Seq ID No
27	82	100.0	543	2	AAV02345	Aay02345 A human h
28	82	100.0	543	2	AAV17082	Aay17082 Human hep
29	82	100.0	543	3	AAV57590	Aay57590 Human hep
30	82	100.0	543	3	AAV08849	Aab08849 Amino aci
31	82	100.0	543	3	AAV52990	Aay52990 Human hep
32	82	100.0	543	4	AAV97635	Aay97635 Human hep
33	82	100.0	543	4	AAV86206	Aab86206 Human hep
34	82	100.0	543	5	ABB07813	Abb07813 Human hep
35	82	100.0	543	7	ADD18950	Adb18950 Human dis
36	82	100.0	543	7	ADG88800	Adg88800 Human hpa
37	82	100.0	543	8	ADL16379	Adl16379 Human hep
38	82	100.0	543	8	ADK52086	Adk52086 Human ato
39	82	100.0	543	8	ADM48716	Adm48716 Human hpa
40	82	100.0	543	8	ADM48759	Adm48759 Human hpa
41	82	100.0	543	8	ADN05074	Adn05074 Antipgori
42	82	100.0	543	8	ADN04902	Adn04902 Antipgori
43	82	100.0	543	8	ADO63831	Ado63831 Human hep
44	82	100.0	543	8	ADO63832	Ado63832 Human hep
45	82	100.0	543	8	ADO63822	Ado63822 Human hep
46	82	100.0	543	8	ADQ80372	Adq80372 Heparanas
47	82	100.0	543	8	ADR88210	Adr88210 Human pre
48	82	100.0	543	8	ADP25079	Adp25079 PRO polyp
49	82	100.0	543	8	ADT78177	Adt78177 Human hep
50	82	100.0	543	9	ADY27036	Ady27036 Human hep
51	82	100.0	543	9	AEA42466	Aea42466 Human hep
52	82	100.0	543	9	AEA42426	Aea42426 Human hep
53	82	100.0	543	10	ABE96848	Aee96848 Human hep
54	82	100.0	545	7	ABP56822	Abp56822 Human hep
55	82	100.0	545	7	ADL16012	Adl16012 G-coupled
56	82	100.0	545	8	ADL93951	Adl93951 Human G-c
57	82	100.0	556	9	ADZ19010	Adz19010 Heparanas
58	82	100.0	570	9	ADZ19008	Adz19008 Heparanas
59	82	100.0	588	2	AAV30124	Aay30124 A human p
60	82	100.0	592	2	AAV02346	Aay02346 A human h
61	82	100.0	592	3	AA808850	Aab08850 Amino aci
62	82	100.0	592	7	ADG88804	Adg88804 Human SK-
63	82	100.0	592	8	ADL16383	Adl16383 Human hep
64	82	100.0	592	8	ADM48720	Adm48720 Human SK-
65	82	100.0	592	9	AEA42461	Aea42461 Human hep
66	77	93.9	15	9	ADU71089	Adu71089 Human hep
67	77	93.9	19	8	ADO63828	Ado63828 Human hep
68	77	93.9	380	2	AAV17085	Aay17085 Rat hepar
69	77	93.9	380	2	AAV17084	Aay17084 Mouse hep
70	77	93.9	535	3	AB808851	Aab08851 A murine
71	77	93.9	535	5	AB807811	Abb07811 Mouse hep
72	77	93.9	535	7	ADG88834	Adg88834 Mouse hpa
73	77	93.9	535	8	ADL16413	Adl16413 Mouse hep
74	77	93.9	535	8	ADM48750	Adm48750 Mouse hpa
75	77	93.9	535	8	ADR88208	Adr88208 Mouse hep
76	77	93.9	535	8	ADT78175	Adt78175 Mouse hep
77	77	93.9	535	9	ADY27033	Ady27033 Murine he
78	77	93.9	535	9	AEA42424	Aea42424 Mouse hep
79	77	93.9	536	5	AB807812	Abb07812 Rat hepar
80	77	93.9	536	8	ADR88209	Adr88209 Rat hepar
81	77	93.9	536	8	ADT78176	Adt78176 Rat hepar
82	77	93.9	536	9	AEA42425	Aea42425 Rat hepar
83	77	93.9	536	9	ADY27035	Ady27035 Rat hepar
84	77	93.9	543	4	AB888361	Aab88361 Human mem
85	77	93.9	543	9	ADY63087	Ady63087 Human clo
86	76	92.7	527	8	ADO63827	Ado63827 Chimeric
87	76	92.7	527	8	ADO63826	Ado63826 Chimeric
88	76	92.7	543	8	ADO63824	Ado63824 Human hep
89	76	92.7	543	8	ADO63823	Ado63823 Human hep
90	75	91.5	15	9	ADU71068	Adu71068 Human hep
91	73	89.0	523	5	AB807814	Abb07814 Chicken h
92	73	89.0	523	7	ABW02017	Abw02017 Chicken h
93	73	89.0	523	8	ADR88211	Adr88211 Chicken h
94	73	89.0	523	8	ADT78178	Adt78178 Chicken h
95	73	89.0	523	9	ADY27037	Ady27037 Chicken h
96	73	89.0	523	9	AEA42427	Aea42427 Chicken h

97	70	85.4	15	9	ADU70873	AdU70873 Human hep	170	42	51.2	833	8	ADM72219	Human TAS
98	68	82.9	545	9	ADY27034	Ady27034 Bovine hep	171	42	51.2	833	8	ADP24693	PRO polyp
99	64	78.0	15	9	ADU70843	AdU70843 Human hep	172	42	51.2	836	8	Adx68344	Plant ful
100	63	72.0	15	9	ADU71217	AdU71217 Human hep	173	42	51.2	847	8	AdR39751	Human kin
101	59	72.0	15	9	ADU70974	AdU70974 Human hep	174	42	51.2	1027	9	ADY70305	Human bet
102	54	65.9	13	2	RAY34185	Ay34185 Human pre	175	42	51.2	1366	2	AAr71702	Collagen
103	53	64.6	9	9	ADU70701	AdU70701 Human hep	176	42	51.2	1366	3	AAy56801	Human pre
104	53	64.6	15	9	ADU71291	AdU71291 Human hep	177	42	51.2	1366	3	AAy96123	Collagen
105	51	62.2	9	9	ADU70549	AdU70549 Human hep	178	42	51.2	1366	4	ABBS0293	Collagen
106	51	62.2	488	4	AAAB31471	AaB31471 Amino aci	179	42	51.2	1366	5	AAE16476	Human col
107	51	62.2	488	4	AAAB31469	AaB31469 Amino aci	180	42	51.2	1366	5	ABB90751	Human Tum
108	50	61.0	9	9	ADU70474	AdU70474 Human hep	181	42	51.2	1366	5	ABB90766	Human Tum
109	49	59.8	9	9	ADU70739	AdU70739 Human hep	182	42	51.2	1366	5	ABB90741	Human Tum
110	48	58.3	9	9	ADU70550	AdU70550 Human hep	183	42	51.2	1366	5	ABB80734	Protein s
111	47	57.3	9	9	ADU70473	AdU70473 Human hep	184	42	51.2	1366	5	ABB09626	Amino aci
112	47	57.3	9	9	ADU70418	AdU70418 Human hep	185	42	51.2	1366	6	ABU54448	Human tum
113	47	57.3	15	9	ADU70973	AdU70973 Human hep	186	42	51.2	1366	6	ABU54473	Human tum
114	47	57.3	488	4	AAAB31470	AaB31470 Amino aci	187	42	51.2	1366	6	ABU54458	Human tum
115	46	56.1	561	6	ABG21696	Abg21696 Protein e	188	42	51.2	1366	6	ABR92065	Human cer
116	46	56.1	639	4	ABG21490	Abg21490 Novel hum	189	42	51.2	1366	7	ADF13076	Human pro
117	46	56.1	1180	6	ABU28742	Abu28742 Protein e	190	42	51.2	1366	8	ADK70437	Respirato
118	45	54.9	134	2	AAr72393	Aar72393 Agmenellu	191	42	51.2	1366	8	ADQ29669	Human col
119	45	54.9	177	8	ADS73853	AdS73853 A. thalia	192	42	51.2	1366	8	ADR16801	Human col
120	45	54.9	523	6	ABU45111	Abu45111 Protein e	193	42	51.2	1366	8	ADR16426	Human col
121	44	53.7	87	4	ABG29684	Abg29684 Novel hum	194	42	51.2	1366	8	ADR99147	Collagen,
122	44	53.7	93	5	ADK36889	AdK36889 Novel hum	195	42	51.2	1366	9	ADV87001	Collagen
123	44	53.7	104	4	AAU48166	Aau48166 Propionib	196	42	51.2	1366	9	ADV70234	Tumor-ass
124	44	53.7	104	6	ABM44685	Abm44685 Propionib	197	42	51.2	1366	9	ADZ70510	Human pro
125	44	53.7	177	4	ABG24421	Abg24421 Novel hum	198	42	51.2	1366	9	ADZ70624	Human pro
126	44	53.7	336	6	ADB23092	AdB23092 Environme	199	42	51.2	1366	9	AEA04491	Human pro
127	44	53.7	341	8	ADQ65272	Adq65272 Novel hum	200	42	51.2	1366	9	AED74603	Human pla
128	44	53.7	354	4	ABG10413	Abg10413 Novel hum	201	42	51.2	1366	10	AEF69981	Colorecta
129	44	53.7	488	4	AAAB31472	AaB31472 Amino aci	202	41	50.0	68	2	AAy31469	N. tabacu
130	44	53.7	563	4	AAU34625	Aau34625 B. coli c	203	41	50.0	68	3	AAb25828	AP2 domai
131	44	53.7	563	6	ABU28684	Abu28684 Protein e	204	41	50.0	91	5	AAm50384	Rat hepar
132	44	53.7	1167	6	ABU27927	Abu27927 Protein e	205	41	50.0	168	4	AAu58491	Propionib
133	43	52.4	9	9	ADU70581	AdU70581 Human hep	206	41	50.0	168	6	ABM55010	Propionib
134	43	52.4	15	9	ADU70872	AdU70872 Human hep	207	41	50.0	172	4	AAU65751	Propionib
135	43	52.4	349	4	AAU50487	Aau50487 Propionib	208	41	50.0	172	6	ABM62270	Propionib
136	43	52.4	349	6	ABM47006	Abm47006 Propionib	209	41	50.0	192	8	ADX91766	Plant ful
137	42	51.2	58	4	AAO07208	Aao07208 Human pol	210	41	50.0	222	8	ADX87478	Plant ful
138	42	51.2	65	4	AAU32257	Aau32257 Novel hum	211	41	50.0	225	5	AAW47870	Tobacco p
139	42	51.2	83	4	AAO09648	Aao09648 Human pol	212	41	50.0	225	7	ADD15139	Tobacco e
140	42	51.2	114	4	AAO13397	Aao13397 Human pol	213	41	50.0	266	10	AEF11596	Soybean m
141	42	51.2	265	7	AAE38636	Aae38636 Human col	214	41	50.0	273	8	ADX65843	Plant ful
142	42	51.2	283	3	AAAG44159	Aag44159 Arabidops	215	41	50.0	281	8	ADT58907	Plant pol
143	42	51.2	283	3	AAAG43083	Aag43083 Arabidops	216	41	50.0	281	10	AEF23235	Lead Cere
144	42	51.2	305	4	AAAG73640	Aag73640 Human col	217	41	50.0	282	8	ADX70963	Plant ful
145	42	51.2	340	9	ADW17545	Adw17545 Pinus rad	218	41	50.0	304	8	ADX68795	Plant ful
146	42	51.2	341	3	AAAG44158	Aag44158 Arabidops	219	41	50.0	312	10	AEF11656	Rice meth
147	42	51.2	341	3	AAAG43082	Aag43082 Arabidops	220	41	50.0	326	10	AEF11652	Rice meth
148	42	51.2	341	8	ADS73843	AdS73843 A. thalia	221	41	50.0	331	5	AAW50383	Human hep
149	42	51.2	341	9	AED60915	Aed60915 Thale cre	222	41	50.0	332	10	AEF11638	Wheat met
150	42	51.2	341	10	AEF26570	Aef26570 A. thalia	223	41	50.0	349	10	AEF11642	Wheat met
151	42	51.2	354	9	ABU17546	Abu17546 Pinus rad	224	41	50.0	355	8	ADX80019	Plant ful
152	42	51.2	355	6	ABU70801	Abu70801 Human adi	225	41	50.0	355	10	AEF11632	Wheat met
153	42	51.2	366	6	ABU15516	Abu15516 Protein e	226	41	50.0	365	5	ABG60934	Novel flo
154	42	51.2	385	5	ADK35029	AdK35029 Novel hum	227	41	50.0	365	10	AEF11664	Rice meth
155	42	51.2	385	5	ABO78366	AbO78366 Pseudomon	228	41	50.0	379	6	ABR40913	Deduced a
156	42	51.2	386	6	ABR41621	AbR41621 Human DIT	229	41	50.0	382	10	AEF11562	Soybean m
157	42	51.2	397	7	ADJ68198	AdJ68198 Human hea	230	41	50.0	384	10	ADW13647	Soybean v
158	42	51.2	464	9	ADW17551	AdW17551 Pinus rad	231	41	50.0	384	10	AEF11630	Soybean m
159	42	51.2	487	7	ABO62870	AbO62870 Klebsiell	232	41	50.0	392	10	AEF11560	Soybean m
160	42	51.2	495	7	ABE37566	AbB37566 Neural th	233	41	50.0	439	4	AAU07423	Human hep
161	42	51.2	495	7	ADJ69822	AdJ69822 Human hea	234	41	50.0	445	7	ADF59424	Human pol
162	42	51.2	534	2	AAW12844	Aaw12844 Pro-alpha	235	41	50.0	470	5	AAE18328	Human hep
163	42	51.2	535	2	AAW12841	Aaw12841 Truncated	236	41	50.0	475	8	ADS23751	Bacterial
164	42	51.2	623	2	AAW12843	Aaw12843 Pro-alpha	237	41	50.0	480	4	AAy97634	Human hep
165	42	51.2	674	3	ABBS3439	AbB3439 Human col	238	41	50.0	480	4	AAU07418	Novel hum
166	42	51.2	740	8	ADR87614	AdR87614 Human typ	239	41	50.0	480	4	AAU07418	Novel hum
167	42	51.2	772	4	ABG01438	AbG01438 Novel hum	240	41	50.0	492	4	AAB84664	Amino aci
168	42	51.2	809	2	AAy29672	Aay29672 Human cer	241	41	50.0	528	5	AAE18327	Human hep
169	42	51.2	833	6	AAO30330	Aao30330 Human MAP	242	41	50.0	534	4	AAAB85216	Heparanas



243	41	50.0	534	5	ABP69310	Abp69310 Human pol	316	40	48.8	4928	2	AAV39300	Aay39300 Spnd a po
244	41	50.0	534	5	AAW50337	Aam50337 Human pre	317	40	48.8	4933	6	ABP57681	Abp57681 Saccharop
245	41	50.0	538	4	AAV97633	Aay97633 Human hep	318	40	48.8	5588	2	AAV39301	Aay39301 Spne a po
246	41	50.0	582	5	AAE18326	Aae18326 Human hep	319	40	48.8	5588	4	AAB70969	Aab70969 S. spinos
247	41	50.0	592	4	AAV97632	Aay97632 Human hep	320	39.5	48.2	171	9	AED71520	Aed71520 Corynebac
248	41	50.0	592	4	AAAB1062	Aab81062 Human Hep	321	39.5	48.2	383	4	AAG93067	Aag93067 C glutami
249	41	50.0	592	4	AAU07424	Aau07424 Human hep	322	39.5	48.2	385	7	ABO65564	Abg65564 Klebsiell
250	41	50.0	592	4	AAAB85215	Aab85215 Heparanas	323	39	47.6	9	ADU70769	Adu70769 Human hep	
251	41	50.0	1130	8	ADT89111	Adt89111 Soybean M	324	39	47.6	15	ADU71070	Adu71070 Human hep	
252	40	48.8	63	4	AAW23516	Aam23516 Human EST	325	39	47.6	30	ADH83892	Adh83892 Bacillus	
253	40	48.8	68	2	AAW23551	Aam23551 Human EST	326	39	47.6	62	AEO08825	Aeo08825 Ovule dev	
254	40	48.8	68	2	AAV31473	Aay31473 A. thalia	327	39	47.6	84	AAO09797	Aao09797 Human pol	
255	40	48.8	68	3	ABE25832	Aab25832 AP2 domai	328	39	47.6	91	AAO61040	Aap61040 N-termina	
256	40	48.8	69	2	AAV31470	Aay31470 N. tabacu	329	39	47.6	100	3	AAO33003	Aao33003 Pinus rad
257	40	48.8	69	3	AAE25829	Aab25829 AP2 domai	330	39	47.6	103	4	AAO12779	Aao12779 Human pol
258	40	48.8	103	7	ABO63688	Aboc63688 Klebsiell	331	39	47.6	105	5	ABP11307	Abp11307 Human ORF
259	40	48.8	109	3	AAV19715	Aag19715 Arabidops	332	39	47.6	127	4	AAO12263	Aao12263 Human pol
260	40	48.8	112	3	AAU19714	Aau19714 Arabidops	333	39	47.6	128	9	AEA26323	Aea26323 Stress to
261	40	48.8	119	4	AAU02524	Aau02524 Anti-adip	334	39	47.6	128	9	AEA27246	Aea27246 Stress to
262	40	48.8	151	9	ADW18623	Adw18623 Eucalyptu	335	39	47.6	128	9	AEF28690	Aef28690 Lead_Cere
263	40	48.8	167	3	ABP32789	Aab32789 Eucalyptu	336	39	47.6	130	4	AAW17187	Aam17187 Peptide #
264	40	48.8	193	10	AEF11612	Aef11612 Soybean m	337	39	47.6	130	4	ABB36190	Abb36190 Peptide #
265	40	48.8	193	10	AEF11602	Aef11602 Soybean m	338	39	47.6	130	4	AAW29682	Aam29682 Peptide #
266	40	48.8	198	8	ADX70784	Adx70784 Plant ful	339	39	47.6	130	4	ABB30994	Abb30994 Peptide #
267	40	48.8	200	7	ADN30468	Adn30468 Plant yie	340	39	47.6	130	4	ABB21565	Abb21565 Protein #
268	40	48.8	200	8	ADI43757	Adi43757 Plant tra	341	39	47.6	130	4	AAW69355	Aam69355 Human bon
269	40	48.8	204	8	ADX70856	Adx70856 Plant ful	342	39	47.6	130	4	AAW56968	Aam56968 Human bra
270	40	48.8	210	4	ABG13427	Abg13427 Novel hum	343	39	47.6	130	4	ABG51031	Abg51031 Human liv
271	40	48.8	210	8	ADX70847	Adx70847 Plant ful	344	39	47.6	130	4	AAW04884	Aam04884 Peptide #
272	40	48.8	213	4	AAU54490	Aau54490 Propionib	345	39	47.6	130	5	ABG38974	Abg38974 Human pep
273	40	48.8	213	6	ABW51009	Abw51009 Propionib	346	39	47.6	130	8	ADG63744	Adg63744 Transcript
274	40	48.8	220	8	ADX71560	Adx71560 Plant ful	347	39	47.6	130	9	AEA26237	Aea26237 Stress to
275	40	48.8	222	5	AAU93108	Aau93108 Arabidops	348	39	47.6	130	9	AEA27248	Aea27248 Stress to
276	40	48.8	222	7	ADD55844	Add55844 Thalecres	349	39	47.6	156	2	AAR10439	Aar10439 Prepro su
277	40	48.8	222	8	ADN01595	Adn01595 Thalecres	350	39	47.6	158	8	ADX66520	Adx66520 Plant ful
278	40	48.8	222	8	ADN73293	Adn73293 Thale cre	351	39	47.6	166	7	ABM86570	Abm86570 Rice abio
279	40	48.8	222	9	ADW17092	Adw17092 Eucalyptu	352	39	47.6	166	10	AEF28702	Aef28702 Lead_Cere
280	40	48.8	229	7	ADC84613	Adc84613 Chandipur	353	39	47.6	171	4	AAW78376	Aam78376 Human pro
281	40	48.8	232	4	AAE08544	Aae08544 Phase M10	354	39	47.6	172	4	AAW00037	Aam00037 Beta-1, 3
282	40	48.8	232	8	ADG65815	Adg65815 Plant ful	355	39	47.6	173	4	ABG04008	Abg04008 Novel hum
283	40	48.8	261	7	ABO75261	Abg75261 Pseudomon	356	39	47.6	175	8	ADP46537	Adp46537 Human col
284	40	48.8	264	5	ABW04661	Abw04661 A. thalia	357	39	47.6	184	8	ADI61313	Adi61313 A. thalia
285	40	48.8	286	9	ADW17105	Adw17105 Eucalyptu	358	39	47.6	184	8	ADO02321	Ado02321 Thalecres
286	40	48.8	303	8	ADX87832	Adx87832 Plant ful	359	39	47.6	184	9	AEA26599	Aea26599 Stress to
287	40	48.8	317	9	ADW17108	Adw17108 Eucalyptu	360	39	47.6	194	7	ABM86302	Abm86302 Rice abio
288	40	48.8	328	9	ADW18206	Adw18206 E. grandis	361	39	47.6	196	3	AAW16018	Aaw16018 E. coli p
289	40	48.8	366	6	ABR55427	Abw55427 Amino aci	362	39	47.6	196	4	AAG98320	Aag98320 Escherich
290	40	48.8	366	10	AEF11588	Aef11588 Tomato me	363	39	47.6	198	7	ABM90275	Abm90275 Rice abio
291	40	48.8	369	8	ADU73566	Adu73566 Hot peppe	364	39	47.6	198	7	ABM90061	Abm90061 Rice abio
292	40	48.8	371	7	ADK72630	Adk72630 AP2 DNA b	365	39	47.6	203	10	AEF28699	Aef28699 Lead_Cere
293	40	48.8	372	10	AEF11648	Aef11648 Tomato me	366	39	47.6	203	10	AEF28700	Aef28700 Lead_Cere
294	40	48.8	379	8	ADG90615	Adg90615 Plant ful	367	39	47.6	205	10	AEF11614	Aef11614 Soybean m
295	40	48.8	493	4	ABW65502	Abw65502 Drosophil	368	39	47.6	207	8	ABO59992	Abg59992 Human gen
296	40	48.8	539	4	ABG04950	Abg04950 Novel hum	369	39	47.6	208	8	ADY05456	Ady05456 Plant ful
297	40	48.8	582	4	ABG04945	Abg04945 Novel hum	370	39	47.6	208	10	AEF29342	Aef29342 Lead_Cere
298	40	48.8	582	7	ABM88330	Abm88330 Rice abio	371	39	47.6	211	1	AAP94616	Aap94616 Alkaline
299	40	48.8	584	8	ADY09948	Ady09948 Plant ful	372	39	47.6	225	6	AAE38264	Aae38264 Rice dise
300	40	48.8	599	4	ABG18132	Abg18132 Novel hum	373	39	47.6	225	8	ADX68974	Adx68974 Plant ful
301	40	48.8	636	8	ADX73228	Adx73228 Plant ful	374	39	47.6	231	9	AED34962	Aed34962 Nicotiana
302	40	48.8	723	9	ABM94228	Abm94228 M. xanthu	375	39	47.6	240	10	AEF28701	Aef28701 Lead_Cere
303	40	48.8	723	9	AEC08348	Aec08348 A. Evansi	376	39	47.6	240	10	AEF28165	Aef28165 Lead_Cere
304	40	48.8	773	9	AEC08348	Aec08348 A. Evansi	377	39	47.6	241	10	AEF11546	Aef11546 Rape meth
305	40	48.8	797	4	ABG18136	Abg18136 Novel hum	378	39	47.6	241	10	AEF11532	Aef11532 Rape meth
306	40	48.8	823	4	ABG09263	Abg09263 Novel hum	379	39	47.6	244	10	AEF11538	Aef11538 Rape meth
307	40	48.8	1128	2	AAW05107	Aaw05107 Sequence	380	39	47.6	245	10	AEF28164	Aef28164 Lead_Cere
308	40	48.8	1666	7	ADG71666	Adg71666 Chlamydom	381	39	47.6	262	8	ADX73070	Adx73070 Plant ful
309	40	48.8	1779	8	ADH56614	Adh56614 Deduced p	382	39	47.6	265	2	AAR10011	Aar10011 Type 4 GX
310	40	48.8	1856	3	AAV21802	Aav21802 B. subtil	383	39	47.6	265	2	AAW53294	Aaw53294 IGG-bindi
311	40	48.8	1856	3	AAV83270	Aay83270 Polypepti	384	39	47.6	273	8	ADO02598	Ado02598 Thalecres
312	40	48.8	2152	2	AAV39298	Aav39298 SpnB a po	385	39	47.6	273	8	ADO62233	Ado62233 Transcript
313	40	48.8	2152	4	AAV70966	Aab70966 S. spinos	386	39	47.6	279	6	ABU40429	Abu40429 Protein e
314	40	48.8	3295	8	ADH39702	Adh39702 Streptomy	387	39	47.6	279	7	ADH86062	Adh86062 Enterococ
315	40	48.8	3651	8	ADH39704	Adh39704 Streptomy	388	39	47.6	279	7	ADH86062	Adh86062 Enterococ
316	40	48.8	4924	4	AAW70968	Aab70968 S. spinos	388	39	47.6	280	6	ABU41977	Abu41977 Protein e

389	39	47.6	315	9	ADW17543	Adw17543 Pinus rad	462	39	47.6	382	2	AAR74223	B. amylololi
390	39	47.6	327	10	AEF11566	Aef11566 Tomato me	463	39	47.6	382	2	AAR74223	B. amylololi
391	39	47.6	332	9	ABM93316	Abm93316 M. xanthu	464	39	47.6	382	2	AAW00247	Subtilisi
392	39	47.6	342	4	AAU49617	Aau49617 Propionib	465	39	47.6	382	2	AAW00247	Subtilisi
393	39	47.6	342	6	ADM46136	Adm46136 Propionib	466	39	47.6	382	2	AAW00247	Subtilisi
394	39	47.6	359	9	ADW19790	Adw19790 Pinus rad	467	39	47.6	382	2	AAR96237	Wild type
395	39	47.6	366	8	ADN18776	Adn18776 Bacterial	468	39	47.6	382	2	AAR96237	Wild type
396	39	47.6	372	8	ADP84789	Adp84789 Glycoside	469	39	47.6	382	2	AAW34776	Bacillus
397	39	47.6	373	2	AAR86878	Aar86878 Subtilisi	470	39	47.6	382	2	AAW34776	Bacillus
398	39	47.6	373	2	AAR86877	Aar86877 Subtilisi	471	39	47.6	382	2	AAW34776	Bacillus
399	39	47.6	373	2	AAR86879	Aar86879 Subtilisi	472	39	47.6	382	2	AAW34776	Bacillus
400	39	47.6	373	2	AAR96240	Aar96240 Mutant su	473	39	47.6	382	2	AAW34776	Bacillus
401	39	47.6	373	2	AAR96244	Aar96244 Mutant su	474	39	47.6	382	2	AAW34776	Bacillus
402	39	47.6	373	2	AAR96245	Aar96245 Mutant su	475	39	47.6	382	2	AAW34776	Bacillus
403	39	47.6	373	2	AAR96242	Aar96242 Mutant su	476	39	47.6	382	2	AAW34776	Bacillus
404	39	47.6	373	2	AAR96238	Aar96238 Mutant su	477	39	47.6	382	2	AAW34776	Bacillus
405	39	47.6	373	2	AAR96241	Aar96241 Mutant su	478	39	47.6	382	2	AAW34776	Bacillus
406	39	47.6	373	2	AAR96243	Aar96243 Mutant su	479	39	47.6	382	2	AAW34776	Bacillus
407	39	47.6	373	2	AAR96239	Aar96239 Mutant su	480	39	47.6	382	2	AAW34776	Bacillus
408	39	47.6	373	2	AAW41692	Aaw41692 BPN' subs	481	39	47.6	382	2	AAW34776	Bacillus
409	39	47.6	373	5	ABB79110	Abb79110 Mutant su	482	39	47.6	382	2	AAW34776	Bacillus
410	39	47.6	373	5	ABB79092	Abb79092 Mutant su	483	39	47.6	382	2	AAW34776	Bacillus
411	39	47.6	373	5	ABB79094	Abb79094 Mutant su	484	39	47.6	382	2	AAW34776	Bacillus
412	39	47.6	373	5	ABB79101	Abb79101 Mutant su	485	39	47.6	382	2	AAW34776	Bacillus
413	39	47.6	373	5	ABB79107	Abb79107 Mutant su	486	39	47.6	382	2	AAW34776	Bacillus
414	39	47.6	373	5	ABB79100	Abb79100 Mutant su	487	39	47.6	382	2	AAW34776	Bacillus
415	39	47.6	373	5	ABB79108	Abb79108 Mutant su	488	39	47.6	382	2	AAW34776	Bacillus
416	39	47.6	373	5	ABB79109	Abb79109 Mutant su	489	39	47.6	382	2	AAW34776	Bacillus
417	39	47.6	373	5	ABB79093	Abb79093 Mutant su	490	39	47.6	382	2	AAW34776	Bacillus
418	39	47.6	373	5	ABB79117	Abb79117 Mutant su	491	39	47.6	382	2	AAW34776	Bacillus
419	39	47.6	373	5	ABB79123	Abb79123 Mutant su	492	39	47.6	382	2	AAW34776	Bacillus
420	39	47.6	373	5	ABB79097	Abb79097 Mutant su	493	39	47.6	382	2	AAW34776	Bacillus
421	39	47.6	373	5	ABB79120	Abb79120 Mutant su	494	39	47.6	382	2	AAW34776	Bacillus
422	39	47.6	373	5	ABB79091	Abb79091 Mutant su	495	39	47.6	382	2	AAW34776	Bacillus
423	39	47.6	373	5	ABB79102	Abb79102 Mutant su	496	39	47.6	382	2	AAW34776	Bacillus
424	39	47.6	373	5	ABB79115	Abb79115 Mutant su	497	39	47.6	382	2	AAW34776	Bacillus
425	39	47.6	373	5	ABB79090	Abb79090 Mutant su	498	39	47.6	382	2	AAW34776	Bacillus
426	39	47.6	373	5	ABB79113	Abb79113 Mutant su	499	39	47.6	382	2	AAW34776	Bacillus
427	39	47.6	373	5	ABB79121	Abb79121 Mutant su	500	39	47.6	382	2	AAW34776	Bacillus
428	39	47.6	373	5	ABB79099	Abb79099 Mutant su	501	39	47.6	382	2	AAW34776	Bacillus
429	39	47.6	373	5	ABB79111	Abb79111 Mutant su	502	39	47.6	382	2	AAW34776	Bacillus
430	39	47.6	373	5	ABB79122	Abb79122 Mutant su	503	39	47.6	382	2	AAW34776	Bacillus
431	39	47.6	373	5	ABB79125	Abb79125 Mutant su	504	39	47.6	382	2	AAW34776	Bacillus
432	39	47.6	373	5	ABB79095	Abb79095 Mutant su	505	39	47.6	382	2	AAW34776	Bacillus
433	39	47.6	373	5	ABB79103	Abb79103 Mutant su	506	39	47.6	382	2	AAW34776	Bacillus
434	39	47.6	373	5	ABB79104	Abb79104 Mutant su	507	39	47.6	382	2	AAW34776	Bacillus
435	39	47.6	373	5	ABB79119	Abb79119 Mutant su	508	39	47.6	382	2	AAW34776	Bacillus
436	39	47.6	373	5	ABB79124	Abb79124 Mutant su	509	39	47.6	382	2	AAW34776	Bacillus
437	39	47.6	373	5	ABB79098	Abb79098 Mutant su	510	39	47.6	382	2	AAW34776	Bacillus
438	39	47.6	373	5	ABB79114	Abb79114 Mutant su	511	39	47.6	382	2	AAW34776	Bacillus
439	39	47.6	373	5	ABB79116	Abb79116 Mutant su	512	39	47.6	382	2	AAW34776	Bacillus
440	39	47.6	373	5	ABB79096	Abb79096 Mutant su	513	39	47.6	382	2	AAW34776	Bacillus
441	39	47.6	373	5	ABB79112	Abb79112 Mutant su	514	39	47.6	382	2	AAW34776	Bacillus
442	39	47.6	373	5	ABB79089	Abb79089 Mutant su	515	39	47.6	382	2	AAW34776	Bacillus
443	39	47.6	373	5	ABB79105	Abb79105 Mutant su	516	39	47.6	382	2	AAW34776	Bacillus
444	39	47.6	373	5	ABB79118	Abb79118 Mutant su	517	39	47.6	382	2	AAW34776	Bacillus
445	39	47.6	373	5	ABB79110	Abb79110 Mutant su	518	39	47.6	382	2	AAW34776	Bacillus
446	39	47.6	373	5	ABB79117	Abb79117 Mutant su	519	39	47.6	382	2	AAW34776	Bacillus
447	39	47.6	373	5	ABB79124	Abb79124 Mutant su	520	39	47.6	382	2	AAW34776	Bacillus
448	39	47.6	373	5	ABB79098	Abb79098 Mutant su	521	39	47.6	382	2	AAW34776	Bacillus
449	39	47.6	373	5	ABB79114	Abb79114 Mutant su	522	39	47.6	382	2	AAW34776	Bacillus
450	39	47.6	373	5	ABB79116	Abb79116 Mutant su	523	39	47.6	382	2	AAW34776	Bacillus
451	39	47.6	373	5	ABB79096	Abb79096 Mutant su	524	39	47.6	382	2	AAW34776	Bacillus
452	39	47.6	373	5	ABB79112	Abb79112 Mutant su	525	39	47.6	382	2	AAW34776	Bacillus
453	39	47.6	373	5	ABB79089	Abb79089 Mutant su	526	39	47.6	382	2	AAW34776	Bacillus
454	39	47.6	373	5	ABB79105	Abb79105 Mutant su	527	39	47.6	382	2	AAW34776	Bacillus
455	39	47.6	373	5	ABB79118	Abb79118 Mutant su	528	39	47.6	382	2	AAW34776	Bacillus
456	39	47.6	373	5	ABB79110	Abb79110 Mutant su	529	39	47.6	382	2	AAW34776	Bacillus
457	39	47.6	373	5	ABB79117	Abb79117 Mutant su	530	39	47.6	382	2	AAW34776	Bacillus
458	39	47.6	373	5	ABB79124	Abb79124 Mutant su	531	39	47.6	382	2	AAW34776	Bacillus
459	39	47.6	373	5	ABB79098	Abb79098 Mutant su	532	39	47.6	382	2	AAW34776	Bacillus
460	39	47.6	373	5	ABB79114	Abb79114 Mutant su	533	39	47.6	382	2	AAW34776	Bacillus
461	39	47.6	373	5	ABB79116	Abb79116 Mutant su	534	39	47.6	382	2	AAW34776	Bacillus

535	39	47.6	744	9	ADX26414	Novel cel	608	38	46.3	189	8	ADQ65251	Novel hum
536	39	47.6	744	9	ADX26271	Novel cel	609	38	46.3	192	4	ABA48292	Human ZF4
537	39	47.6	751	4	ABG04583	Novel hum	610	38	46.3	193	3	AGS60567	Arabidops
538	39	47.6	751	4	ABG16291	Novel hum	611	38	46.3	197	8	ADW18204	E.grandis
539	39	47.6	762	4	AAW79390	Human pro	612	38	46.3	201	9	ADW18204	E.grandis
540	39	47.6	764	8	ADR66089	Human pro	613	38	46.3	202	8	ADT60371	Plant pol
541	39	47.6	764	8	ADR66431	Human pro	614	38	46.3	210	2	AAW31709	X25 gene
542	39	47.6	846	8	ADO61707	Transcrip	615	38	46.3	210	3	AAW31709	X25 gene
543	39	47.6	865	9	AED07619	Aerononas	616	38	46.3	210	7	ADJ69448	Human hea
544	39	47.6	866	2	AAW02159	Soluble c	617	38	46.3	216	8	ADT58219	Plant pol
545	39	47.6	866	3	AAW52307	Vibrio fu	618	38	46.3	216	8	ADX69005	Plant ful
546	39	47.6	879	6	ABJ25894	Aspergill	619	38	46.3	218	3	AGS05951	Arabidops
547	39	47.6	988	4	ABG16888	Novel hum	620	38	46.3	218	3	AGS05951	Arabidops
548	39	47.6	988	7	ADI21719	Novel hum	621	38	46.3	218	3	AGS05951	Arabidops
549	39	47.6	1032	6	ABJ26494	Aspergill	622	38	46.3	218	5	AAU33058	Arabidops
550	39	47.6	1366	4	AAE02536	Aspergill	623	38	46.3	218	5	AAU33058	Arabidops
551	39	47.6	1372	4	ADT59683	Forcine a	624	38	46.3	218	7	ADD30029	Plant yie
552	39	47.6	1372	7	ADT59683	Rat Prote	625	38	46.3	218	7	ADD30029	Plant yie
553	39	47.6	1372	7	ADD45148	Rat Prote	626	38	46.3	218	7	ADD30029	Plant yie
554	39	47.6	1372	7	ADD45604	Rat Prote	627	38	46.3	218	7	ADD30029	Plant yie
555	39	47.6	1372	7	ADT59687	Rat Prote	628	38	46.3	218	8	ADT59687	Rat Prote
556	39	47.6	1373	5	ABJ57364	Mouse isc	629	38	46.3	218	9	ADT59687	Rat Prote
557	39	47.6	1373	9	ADW44460	Murine pr	630	38	46.3	218	9	ADT59687	Rat Prote
558	39	47.6	1373	10	AEF119250	Mus muscu	631	38	46.3	218	9	AEF119250	Mus muscu
559	38.5	47.0	142	2	AAE50193	Heavy cha	632	38	46.3	218	9	AEF119250	Mus muscu
560	38.5	47.0	142	2	AAE50193	Heavy cha	633	38	46.3	218	10	AEF119250	Mus muscu
561	38.5	47.0	142	8	ADT87532	Humanised	634	38	46.3	218	10	AEF119250	Mus muscu
562	38.5	47.0	734	9	AAW93705	M. xanthu	635	38	46.3	218	10	AEF119250	Mus muscu
563	38	46.3	9	3	AAW70256	Peptide e	636	38	46.3	227	9	ABM96418	M. xanthu
564	38	46.3	33	10	AEF119250	Human PY	637	38	46.3	227	9	ABM96418	M. xanthu
565	38	46.3	33	10	AEF119250	Human PY	638	38	46.3	227	9	ABM96418	M. xanthu
566	38	46.3	33	10	AEF119250	Human PY	639	38	46.3	227	9	ABM96418	M. xanthu
567	38	46.3	33	10	AEF119250	Human PY	640	38	46.3	227	9	ABM96418	M. xanthu
568	38	46.3	62	4	ADH75729	Thalecres	641	38	46.3	227	9	ABM96418	M. xanthu
569	38	46.3	68	2	AAW31472	A. thalia	642	38	46.3	227	9	ABM96418	M. xanthu
570	38	46.3	68	2	AAW31472	A. thalia	643	38	46.3	227	9	ABM96418	M. xanthu
571	38	46.3	68	3	ABW25834	AP2 domai	644	38	46.3	227	9	ABM96418	M. xanthu
572	38	46.3	68	3	ABW25834	AP2 domai	645	38	46.3	227	9	ABM96418	M. xanthu
573	38	46.3	68	3	ABW25834	AP2 domai	646	38	46.3	227	9	ABM96418	M. xanthu
574	38	46.3	68	3	ABW25834	AP2 domai	647	38	46.3	227	9	ABM96418	M. xanthu
575	38	46.3	102	5	AAO17456	Human liv	648	38	46.3	227	9	ABM96418	M. xanthu
576	38	46.3	124	8	ADL43274	Plant tra	649	38	46.3	227	9	ABM96418	M. xanthu
577	38	46.3	124	8	ADL43274	Plant tra	650	38	46.3	227	9	ABM96418	M. xanthu
578	38	46.3	124	9	ADL43274	Plant tra	651	38	46.3	227	9	ABM96418	M. xanthu
579	38	46.3	124	9	ADL43274	Plant tra	652	38	46.3	227	9	ABM96418	M. xanthu
580	38	46.3	124	10	AEF28687	Lead_Cere	653	38	46.3	227	9	ABM96418	M. xanthu
581	38	46.3	128	10	AEF28687	Lead_Cere	654	38	46.3	227	9	ABM96418	M. xanthu
582	38	46.3	138	8	ADH09769	Human hos	655	38	46.3	227	9	ABM96418	M. xanthu
583	38	46.3	138	8	ADH09769	Human hos	656	38	46.3	227	9	ABM96418	M. xanthu
584	38	46.3	138	9	AEA26245	Stress to	657	38	46.3	227	9	ABM96418	M. xanthu
585	38	46.3	138	9	AEA26245	Stress to	658	38	46.3	227	9	ABM96418	M. xanthu
586	38	46.3	138	9	AEA26245	Stress to	659	38	46.3	227	9	ABM96418	M. xanthu
587	38	46.3	138	10	AEF28687	Lead_Cere	660	38	46.3	227	9	ABM96418	M. xanthu
588	38	46.3	138	10	AEF28687	Lead_Cere	661	38	46.3	227	9	ABM96418	M. xanthu
589	38	46.3	138	10	AEF28687	Lead_Cere	662	38	46.3	227	9	ABM96418	M. xanthu
590	38	46.3	138	10	AEF28687	Lead_Cere	663	38	46.3	227	9	ABM96418	M. xanthu
591	38	46.3	138	10	AEF28687	Lead_Cere	664	38	46.3	227	9	ABM96418	M. xanthu
592	38	46.3	138	10	AEF28687	Lead_Cere	665	38	46.3	227	9	ABM96418	M. xanthu
593	38	46.3	141	3	AGS45282	Arabidops	666	38	46.3	227	9	ABM96418	M. xanthu
594	38	46.3	141	3	AGS45282	Arabidops	667	38	46.3	227	9	ABM96418	M. xanthu
595	38	46.3	141	3	AGS45282	Arabidops	668	38	46.3	227	9	ABM96418	M. xanthu
596	38	46.3	141	3	AGS45282	Arabidops	669	38	46.3	227	9	ABM96418	M. xanthu
597	38	46.3	141	3	AGS45282	Arabidops	670	38	46.3	227	9	ABM96418	M. xanthu
598	38	46.3	141	3	AGS45282	Arabidops	671	38	46.3	227	9	ABM96418	M. xanthu
599	38	46.3	141	3	AGS45282	Arabidops	672	38	46.3	227	9	ABM96418	M. xanthu
600	38	46.3	141	3	AGS45282	Arabidops	673	38	46.3	227	9	ABM96418	M. xanthu
601	38	46.3	141	3	AGS45282	Arabidops	674	38	46.3	227	9	ABM96418	M. xanthu
602	38	46.3	141	3	AGS45282	Arabidops	675	38	46.3	227	9	ABM96418	M. xanthu
603	38	46.3	141	3	AGS45282	Arabidops	676	38	46.3	227	9	ABM96418	M. xanthu
604	38	46.3	141	3	AGS45282	Arabidops	677	38	46.3	227	9	ABM96418	M. xanthu
605	38	46.3	141	3	AGS45282	Arabidops	678	38	46.3	227	9	ABM96418	M. xanthu
606	38	46.3	141	3	AGS45282	Arabidops	679	38	46.3	227	9	ABM96418	M. xanthu
607	38	46.3	141	3	AGS45282	Arabidops	680	38	46.3	227	9	ABM96418	M. xanthu

681	38	46.3	309	9	ADM17107	Adw17107 Eucalyptu	754	38	46.3	750	7	ADM04450	Adm04450 Human pro
682	38	46.3	310	4	AAE03575	Aae03575 Human pro	755	38	46.3	750	9	AEC87380	Aec87380 Human cdn
683	38	46.3	317	4	AAE03579	Aae03579 Human pro	756	38	46.3	757	4	AEC87383	Aae03583 Human pro
684	38	46.3	329	10	AEF11568	Aef11568 Coffee me	757	38	46.3	810	7	ADE62643	Ade62643 Rat Prote
685	38	46.3	331	9	AEA57858	Aea57858 Novel S.	758	38	46.3	810	7	ADE62647	Ade62647 Rat Prote
686	38	46.3	331	9	AEA57858	Aea57858 Streptoco	759	38	46.3	810	8	ADR32025	Adr32025 Rat NELL1
687	38	46.3	333	3	AG227694	Aag227694 Arabidops	760	38	46.3	852	6	ABU12116	Abu12116 Human pro
688	38	46.3	335	3	AA037354	Aa037354 Dehydrati	761	38	46.3	879	4	ABG25755	Abg25755 Novel hum
689	38	46.3	335	3	AA037354	Aa037354 Dehydrati	762	38	46.3	879	4	ABG28383	Abg28383 Novel hum
690	38	46.3	335	3	AA037354	Aa037354 Dehydrati	763	38	46.3	908	4	AAE03572	Aae03572 Human pro
691	38	46.3	335	3	AAE02539	Aae02539 A. thalia	764	38	46.3	909	8	ADS23793	Ads23793 Bacterial
692	38	46.3	335	7	ADD55736	Add55736 Thalecres	765	38	46.3	950	7	AEE72771	Aee72771 Novel hum
693	38	46.3	335	7	ADD30754	Add30754 Plant yie	766	38	46.3	959	4	AB86947	Ab86947 Human met
694	38	46.3	335	8	ADI43763	Adi43763 Plant tra	767	38	46.3	971	8	ABM81634	Abm81634 Tumour-as
695	38	46.3	335	8	ADO01629	Ado01629 Thalecres	768	38	46.3	972	6	ABU11811	Abu11811 Human MDD
696	38	46.3	335	8	ADP22718	Adp22718 Arabidops	769	38	46.3	972	6	ABU11669	Abu11669 Human MDD
697	38	46.3	335	8	ADS73829	Ads73829 A. thalia	770	38	46.3	1010	8	ADO05302	Ado05302 Chlamydom
698	38	46.3	335	8	ADS19137	Ads19137 Rice DREB	771	38	46.3	1021	8	ADO07041	Ado07041 Human pro
699	38	46.3	335	8	ADT06981	Adt06981 Arabidops	772	38	46.3	1098	7	ADB64321	Adb64321 Human pro
700	38	46.3	335	8	AEA26985	Aea26985 Stress to	773	38	46.3	1117	7	ADZ44782	Adz44782 Human ang
701	38	46.3	335	10	AEF31036	Aef31036 Arabidops	774	38	46.3	1120	5	ABG76502	Abg76502 DNA encod
702	38	46.3	342	7	ADJ11386	Adj11386 Rice prot	775	38	46.3	1124	8	ADT89120	Adt89120 Tomato MS
703	38	46.3	342	7	ADJ11726	Adj11726 Rice prot	776	38	46.3	1212	6	ABU40820	Abu40820 Protein e
704	38	46.3	342	7	ABM85988	Abm85988 Rice abio	777	38	46.3	1213	7	ADF06193	Adf06193 Bacterial
705	38	46.3	342	7	ABM90033	Abm90033 Rice abio	778	38	46.3	1343	8	ADS23976	Ads23976 Bacterial
706	38	46.3	342	10	AEF11628	Aef11628 Rice meth	779	38	46.3	1404	4	ABG14929	Abg14929 Novel hum
707	38	46.3	346	8	ADY06604	Ady06604 Plant ful	780	38	46.3	1558	8	ADT89145	Adt89145 MSH1conse
708	38	46.3	346	8	ADY24678	Ady24678 Plant ful	781	37.5	45.7	124	4	AUS58334	Aus58334 Propionib
709	38	46.3	355	8	ADX68445	Adx68445 Plant ful	782	37.5	45.7	124	6	ABM54853	Abm54853 Propionib
710	38	46.3	356	4	AAE03581	Aae03581 Human pro	783	37.5	45.7	152	6	AAO30394	Aao30394 Rabbit 14
711	38	46.3	358	3	AGS22903	Ags22903 Arabidops	784	37.5	45.7	152	8	ADO79140	Ado79140 Anti-CD83
712	38	46.3	358	3	AGS26238	Ags26238 Arabidops	785	37.5	45.7	290	3	AY95037	Ay95037 Candida a
713	38	46.3	358	7	ADB31779	Adb31779 Plant (A.	786	37.5	45.7	291	10	AEF79950	Aef79950 Growth-/y
714	38	46.3	358	7	ADD55714	Add55714 Thalecres	787	37.5	45.7	454	6	AAO30347	Aao30347 Rabbit 14
715	38	46.3	358	8	ADO01607	Ado01607 Thalecres	788	37.5	45.7	454	6	AAO30347	Aao30347 Rabbit 14
716	38	46.3	358	10	AEF11592	Aef11592 A. thalia	789	37.5	45.7	454	6	AAO30347	Aao30347 Rabbit 14
717	38	46.3	358	10	AEF11644	Aef11644 A. thalia	790	37	45.1	24	4	AAH85520	Aah85520 Heparanas
718	38	46.3	362	8	ADX93368	Adx93368 Plant ful	791	37	45.1	56	8	ABO54957	Abos4957 Human gen
719	38	46.3	362	10	AEF11666	Aef11666 Rice meth	792	37	45.1	65	9	ADV42293	Adv42293 Plant AP2
720	38	46.3	364	8	ADY05458	Ady05458 Plant ful	793	37	45.1	65	9	ADV42315	Adv42315 Plant AP2
721	38	46.3	364	8	ADX95844	Adx95844 Plant ful	794	37	45.1	65	9	ADV42291	Adv42291 Plant AP2
722	38	46.3	364	8	ADX95845	Adx95845 Plant ful	795	37	45.1	65	9	ADV42296	Adv42296 Plant AP2
723	38	46.3	375	10	AEF11654	Aef11654 Beech met	796	37	45.1	65	9	ADV42294	Adv42292 Plant AP2
724	38	46.3	378	8	ADX93676	Adx93676 Plant ful	797	37	45.1	65	9	ADV42294	Adv42292 Plant AP2
725	38	46.3	378	10	AEF11590	Aef11590 Beech met	798	37	45.1	66	4	AAU51977	Aau51977 Propionib
726	38	46.3	379	8	ADS42376	Ads42376 Bacterial	799	37	45.1	66	4	AAU51977	Aau51977 Propionib
727	38	46.3	379	8	AB32761	Ab32761 Eucalyptu	800	37	45.1	66	6	ABM48496	Abm48496 Propionib
728	38	46.3	387	10	AEF11634	Aef11634 Tobacco m	801	37	45.1	68	4	AAU63033	Aau63033 Propionib
729	38	46.3	389	8	ADX93781	Adx93781 Plant ful	802	37	45.1	68	6	ABM59552	Abm59552 Propionib
730	38	46.3	397	6	AAE29815	Aae29815 Streptoco	803	37	45.1	72	5	ABP00912	Abp00912 Human ORF
731	38	46.3	418	8	ADY09711	Ady09711 Plant ful	804	37	45.1	73	3	AB28578	Aab28578 Human SCA
732	38	46.3	421	8	ADY24686	Ady24686 Plant ful	805	37	45.1	75	3	AG00782	Aag00782 Human sec
733	38	46.3	422	8	ADY05453	Ady05453 Plant ful	806	37	45.1	76	8	ADM96968	Adm96968 Human pan
734	38	46.3	438	4	AAE03582	Aae03582 Human pro	807	37	45.1	82	4	AAW92673	Aaw92673 Human dig
735	38	46.3	462	7	AD663510	Ad663510 Rat Prote	808	37	45.1	84	5	ABP31965	Abp31965 Human ORF
736	38	46.3	468	4	AAE03574	Aae03574 Human pro	809	37	45.1	100	8	ADY22704	Ady22704 Plant ful
737	38	46.3	468	7	ABO83691	Abos83691 Pseudomon	810	37	45.1	125	5	ADK34873	Adk34873 Novel hum
738	38	46.3	470	2	AAW55998	Aaw55998 Protein S	811	37	45.1	127	9	AEA27236	Aea27236 Stress to
739	38	46.3	470	9	ADY27582	Ady27582 Neoramini	812	37	45.1	131	5	AAU93095	Aau93095 Arabidops
740	38	46.3	470	9	ADY27580	Ady27580 Neoramini	813	37	45.1	131	7	ADD30336	Add30336 Plant yie
741	38	46.3	470	9	ADY27578	Ady27578 Neoramini	814	37	45.1	131	8	ADI43673	Adi43673 Plant tra
742	38	46.3	471	2	ADH29824	Adh29824 Swinepox	815	37	45.1	131	8	AD143273	Ad143273 Plant tra
743	38	46.3	493	9	ADW17552	Adw17552 Pinus rad	816	37	45.1	131	8	ADO62999	Ado62999 Transcript
744	38	46.3	506	7	ABO78690	Abos78690 Pseudomon	817	37	45.1	131	8	ADO63660	Ado63660 Transcript
745	38	46.3	507	4	AAE03576	Aae03576 Human pro	818	37	45.1	131	9	AEA26195	Aea26195 Stress to
746	38	46.3	531	8	ADO13858	Ado13858 Protein e	819	37	45.1	131	9	AEA26231	Aea26231 Stress to
747	38	46.3	570	3	AAE030056	Aae030056 Arabidops	820	37	45.1	131	9	AEA27241	Aea27241 Stress to
748	38	46.3	589	4	AAE03577	Aae03577 Human pro	821	37	45.1	131	10	AEF28691	Aef28691 Lead Cere
749	38	46.3	603	3	AAE030055	Aae030055 Arabidops	822	37	45.1	131	10	AEF28686	Aef28686 Lead_Cere
750	38	46.3	613	4	ABB62232	Abb62232 Drosophyl	823	37	45.1	131	10	AEF28683	Aef28683 Lead_Cere
751	38	46.3	618	3	AAE030054	Aae030054 Arabidops	824	37	45.1	133	9	AEA27237	Aea27237 Stress to
752	38	46.3	662	2	AAE14585	Aay14585 A.thalian	825	37	45.1	137	6	ABM70003	Abm70003 Photornab
753	38	46.3	662	5	ABB93201	Abb93201 Herbicida	826	37	45.1	140	7	ABO61374	Abos61374 Klebsiell

827	37	45.1	141	9	ABM94687	Am94687 H. xanthu	900	37	45.1	450	7	ADG10792	Adg10792 Human STA
828	37	45.1	148	4	AAO06361	Aao06361 Human pol	901	37	45.1	450	7	ADG10790	Adg10790 Human STA
829	37	45.1	149	9	AEA26319	Aea26319 Stress to	902	37	45.1	451	6	ABU89756	Abu89756 Protein d
830	37	45.1	149	9	AEA27252	Aea27252 Stress to	903	37	45.1	451	6	ABU19526	Abu19526 Protein e
831	37	45.1	151	8	ADO63750	Ado63750 Transcrip	904	37	45.1	463	8	ADX74862	Adx74862 Plant ful
832	37	45.1	151	9	AEA26243	Aea26243 Stress to	905	37	45.1	466	7	ADG33854	Adg33854 Actinomyc
833	37	45.1	151	9	AEA27251	Aea27251 Stress to	906	37	45.1	480	8	ADT58933	Adt58933 Plant pol
834	37	45.1	158	8	ADO63748	Ado63748 Transcrip	907	37	45.1	487	4	AAV72635	Aay72635 Exophiala
835	37	45.1	158	9	AEA27250	Aea27250 Stress to	908	37	45.1	487	5	AAE26385	Aae26385 Exophiala
836	37	45.1	158	9	AEA26241	Aea26241 Stress to	909	37	45.1	487	6	ABG75928	Abg75928 Fumonidin
837	37	45.1	160	4	AAU60281	Aau60281 Propionib	910	37	45.1	487	6	ABU07912	Abu07912 Exophiala
838	37	45.1	160	4	ABM56800	Abm56800 Propionib	911	37	45.1	487	6	ABU62939	Abu62939 E. spinif
839	37	45.1	161	2	AAW37090	Aaw37090 Lycopersi	912	37	45.1	494	8	ADS28840	Ads28840 Bacterial
840	37	45.1	162	6	ABO14768	Abol14768 Novel hum	913	37	45.1	538	7	ADE08404	Ade08404 Novel pro
841	37	45.1	180	8	ADX91956	Adx91956 Plant ful	914	37	45.1	545	9	ADEA21110	Aea21110 Novel hum
842	37	45.1	181	9	ADW17088	Adw17088 Eucalyptu	915	37	45.1	574	9	AED48775	Aed48775 Streptomy
843	37	45.1	191	2	AAR39562	Aar39562 Sequence	916	37	45.1	616	6	ABM72146	Abm72146 Staphyloc
844	37	45.1	196	6	ABR53581	Abr53581 Protein g	917	37	45.1	649	8	ADN23351	Adn23351 Bacterial
845	37	45.1	196	7	ADK64334	Adk64334 Disease c	918	37	45.1	663	8	ADS44672	Ads44672 Bacterial
846	37	45.1	196	9	AEA62715	Aea62715 Mitochond	919	37	45.1	686	7	ADC95375	Adc95375 E. faeciu
847	37	45.1	222	8	ADX68377	Adx68377 Plant ful	920	37	45.1	703	8	ADS15435	Ads15435 HSV-1 pol
848	37	45.1	226	9	AED26774	Aed26774 DNA-PKcs	921	37	45.1	735	2	AAW69761	Aaw69761 Acetobact
849	37	45.1	233	6	ADA36235	Ada36235 Acinetoba	922	37	45.1	739	4	ADR86165	Adr86165 Aspergill
850	37	45.1	234	4	ABG03716	Abg03716 Novel hum	923	37	45.1	781	4	AB558388	Ab558388 Drosophil
851	37	45.1	237	8	ADL42039	Adl42039 Plant tra	924	37	45.1	823	5	ABP65373	Abp65373 Bifidobac
852	37	45.1	237	8	ADO02588	Ado02588 Thalecres	925	37	45.1	903	6	ABU38811	Abu38811 Protein e
853	37	45.1	243	7	ABM90204	Abm90204 Rice abio	926	37	45.1	974	5	ABP35581	Abp35581 Fungal 2B
854	37	45.1	244	9	ADV42264	Adv42264 Plant tra	927	37	45.1	988	7	ABO78500	Ab078500 Pseudomon
855	37	45.1	247	9	ADV42266	Adv42266 Plant tra	928	37	45.1	1005	7	ADE98304	Ade98304 Cancer-li
856	37	45.1	256	3	AGL13479	Agl13479 Arabidops	929	37	45.1	1032	7	ADE98279	Ade98279 Cancer-li
857	37	45.1	259	8	ADL42038	Adl42038 Plant tra	930	37	45.1	1039	7	ADE98283	Ade98283 Cancer-li
858	37	45.1	259	8	ADO02587	Ado02587 Thalecres	931	37	45.1	1040	7	ADE98285	Ade98285 Cancer-li
859	37	45.1	265	7	ABM74280	Abm74280 DNA clone	932	37	45.1	1063	7	ADE98280	Ade98280 Cancer-li
860	37	45.1	267	8	ADX90931	Adx90931 Plant ful	933	37	45.1	1066	7	ADE98302	Ade98302 Cancer-li
861	37	45.1	270	5	ABP53670	Abp53670 Lactococc	934	37	45.1	1089	7	ADE98305	Ade98305 Cancer-li
862	37	45.1	270	9	ADV42236	Adv42236 Plant tra	935	37	45.1	1125	7	ADE98293	Ade98293 Cancer-li
863	37	45.1	271	8	ADM48058	Adm48058 Polypepti	936	37	45.1	1152	7	ADE98288	Ade98288 Cancer-li
864	37	45.1	271	9	ADV42254	Adv42254 Plant tra	937	37	45.1	1159	7	ADE98281	Ade98281 Cancer-li
865	37	45.1	272	9	ADV42256	Adv42256 Plant tra	938	37	45.1	1162	8	ADL16264	Adl16264 Human nuc
866	37	45.1	274	9	ADW17540	Adw17540 Pinus rad	939	37	45.1	1178	6	ABU32204	Abu32204 Protein e
867	37	45.1	276	8	ADM66569	Adm66569 Plant ful	940	37	45.1	1178	7	ABO63427	Ab063427 Klebsiell
868	37	45.1	278	10	AEF11626	Aef11626 Soybean m	941	37	45.1	1186	6	ABR41085	Abr41085 Human MAP
869	37	45.1	279	8	ADX71478	Adx71478 Plant ful	942	37	45.1	1186	7	ADF09579	Adf09579 Human Sec
870	37	45.1	288	4	ABG09003	Abg09003 Novel hum	943	37	45.1	1186	7	ADE98286	Ade98286 Cancer-li
871	37	45.1	288	4	ABG03400	Abg03400 Novel hum	944	37	45.1	1186	8	ADI24519	Adi24519 Human mod
872	37	45.1	294	9	ADV42258	Adv42258 Plant tra	945	37	45.1	1186	8	ADU60281	Adu60281 Housekeep
873	37	45.1	296	5	ABP38345	Abp38345 Staphyloc	946	37	45.1	1186	9	ADX07600	Adx07600 Cyclin-de
874	37	45.1	299	8	ADS21667	Ads21667 Bacterial	947	37	45.1	1186	9	AEEL18644	Aeel18644 Human Pum
875	37	45.1	303	6	AAE33251	Aae33251 Rice DBP1	948	37	45.1	1186	10	AEF34743	Aef34743 Human PUM
876	37	45.1	303	9	ADV42234	Adv42234 Plant tra	949	37	45.1	1186	10	AEF34745	Aef34745 Human exp
877	37	45.1	310	8	ADX95485	Adx95485 Plant ful	950	37	45.1	1186	10	AEF34747	Aef34747 Human exp
878	37	45.1	311	8	ADX74036	Adx74036 Plant ful	951	37	45.1	1187	6	ABR41084	Abr41084 Human MAP
879	37	45.1	323	4	ABB64496	Abb64496 Drosophil	952	37	45.1	1199	9	AED74298	Aed74298 Human pla
880	37	45.1	327	9	ABE39240	Abe39240 L. pneumo	953	37	45.1	1209	7	ADE98292	Ade98292 Cancer-li
881	37	45.1	344	9	AEEO2491	Aee02491 Human her	954	37	45.1	1441	9	AEE74917	Aee74917 pGJA-P/VA
882	37	45.1	344	9	AEEO2625	Aee02625 Human her	955	37	45.1	1561	2	AAW02098	Aaw02098 S. mutans
883	37	45.1	348	9	AEA27258	Aea27258 Stress to	956	37	45.1	1561	6	ABU44355	Abu44355 Protein e
884	37	45.1	352	8	ADS15433	Ads15433 HSV-1 pol	957	37	45.1	1562	9	ABU91501	Abu91501 Microbial
885	37	45.1	352	9	ABE35793	Abe35793 L. pneumo	958	37	45.1	1690	4	ABE61144	Ab61144 Drosophil
886	37	45.1	355	10	AEF11636	Aef11636 Wheat met	959	37	45.1	1690	4	ABE61173	Ab61173 Drosophil
887	37	45.1	363	8	ADS41611	Ads41611 Bacterial	960	37	45.1	1690	9	ADY85283	Ady85283 Drosophil
888	37	45.1	376	8	ADY08195	Ady08195 Plant ful	961	37	45.1	2130	8	AD074677	Ad074677 Streptomy
889	37	45.1	384	6	ABU21232	Abu21232 Protein e	962	37	45.1	3201	4	ABE82214	Abe82214 Polykerid
890	37	45.1	386	2	AAV06416	Aay06416 Aspergill	963	37	45.1	3798	3	AAV58577	Aay58577 Sorangium
891	37	45.1	416	7	ABO80795	Ab080795 Pseudomon	964	36.5	44.5	76	7	ADJ68338	Adj68338 Human hea
892	37	45.1	419	8	ADY06605	Ady06605 Plant ful	965	36.5	44.5	105	4	AAU57181	Aau57181 Propionib
893	37	45.1	442	8	ADN29959	Adn29959 Bacterial	966	36.5	44.5	136	6	ABM53700	Abm53700 Propionib
894	37	45.1	443	8	ADN19253	Adn19253 Bacterial	967	36.5	44.5	136	6	ABM62364	Abm62364 Propionib
895	37	45.1	443	9	ADM43134	Adm43134 Baker's y	968	36.5	44.5	136	6	ABM62364	Abm62364 Propionib
896	37	45.1	447	5	AAE21507	Aae21507 Human gen	969	36.5	44.5	138	4	ABG13594	Abg13594 Novel hum
897	37	45.1	447	5	ABG64971	Abg64971 Human alb	970	36.5	44.5	183	9	ABE36549	Abe36549 L. pneumo
898	37	45.1	447	8	ADL78238	Adl78238 Albumin f	971	36.5	44.5	183	9	ABE39943	Abe39943 L. pneumo
899	37	45.1	450	7	ADG10678	Adg10678 Human STA	972	36.5	44.5	188	3	ABE42212	Aab42212 Human ORF

973 36.5 44.5 250 5 ABP45634 Human Bly  
 974 Adg96461 Single ch  
 975 Aed78514 Human B L  
 976 Abp43699 Interapti  
 977 Abm15877 Mycobacte  
 978 Aeb49952 E. coli p  
 979 Add30286 Plant yie  
 980 Adl44171 Plant tra  
 981 Aau57839 Propionib  
 982 Abm54358 Propionib  
 983 Abg18548 Novel hum  
 984 Adm05629 Human pro  
 985 Adq17383 Human sof  
 986 Aec88559 Human cDN  
 987 Abg21616 Novel hum  
 988 Adt56533 Plant pol  
 989 Aee75860 Streptomy  
 990 Aau43339 Propionib  
 991 Abm40458 Propionib  
 992 Aay59872 Human nor  
 993 Adt90165 Human tum  
 994 Aag75239 Human col  
 995 Aay48484 Human bre  
 996 Aao13377 Human pol  
 997 Aau53913 Propionib  
 998 Abm50432 Propionib  
 999 Aau40108 Propionib  
 1000 Abm36627 Propionib

## ALIGNMENTS

RESULT 1  
 ADL16423  
 ID ADL16423 standard; peptide; 15 AA.  
 XX  
 AC ADL16423;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Human heparanase antigenic peptide.  
 XX  
 KW Human; heparanase; heparanase-dependent cancer; cancer;  
 KW autoimmune reaction; inflammation; immunogen.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2003236215-A1.  
 XX  
 PD 25-DEC-2003.  
 XX  
 PF 09-JUN-2003; 2003US-00456573.  
 XX  
 PR 31-AUG-1998; 98WO-0017954.  
 PR 01-MAR-1999; 99US-00258892.  
 PR 08-NOV-1999; 99US-00435739.  
 XX  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX  
 PI Pecker I, Vlodavsky I, Feinstein E;  
 XX  
 DR WPI; 2004-070610/07.  
 XX  
 XX New antisense oligonucleotide hybridizable with a polynucleotide encoding  
 PT a polypeptide with heparanase activity, useful for treating diseases such  
 PT as cancer and autoimmune disorders.  
 XX  
 XX Example 17; SEQ ID NO 54; 108pp; English.  
 XX  
 XX The invention relates to an antisense oligonucleotide (ASO) comprising a  
 CC polynucleotide or a polynucleotide analogue of at least 10 bases being

CC hybridisable in vivo, under physiological conditions, with a portion of  
 CC a polynucleotide strand encoding a polypeptide having heparanase  
 CC catalytic activity. Also included are a method of in vivo downregulating  
 CC heparanase activity (comprising administering the ASO in vivo), a method  
 CC of treating a subject suffering from a pathological condition  
 CC (characterised by heparanase activity, comprising administering ASO to  
 CC the subject), a pharmaceutical composition comprising the ASO and a  
 CC carrier, an antisense nucleic acid construct (comprising a promoter  
 CC sequence and a polynucleotide sequence directing the synthesis of an  
 CC antisense RNA sequence of at least 10 bases being hybridisable in vivo,  
 CC under physiological conditions, with a polynucleotide strand encoding a  
 CC polypeptide having heparanase catalytic activity), a method of in vivo  
 CC downregulating heparanase activity (comprising administering in vivo the  
 CC antisense nucleic acid construct), a pharmaceutical composition  
 CC comprising the antisense nucleic acid construct and a carrier, and an  
 CC antisense oligonucleotide comprising a polynucleotide or a polynucleotide  
 CC analogue of at least 10 bases being hybridisable in vivo, under  
 CC physiological conditions, with a portion of a polynucleotide strand being  
 CC characterised by forming at least a portion of an untranslated region  
 CC (UTR) for a polynucleotide strand encoding a polypeptide having  
 CC heparanase catalytic activity. The methods and compositions of the  
 CC present invention are useful for the prevention and/or treatment of  
 CC diseases or conditions associated with aberrant heparanase activity, such  
 CC as heparanase-dependent cancer, cancer, autoimmune reaction and  
 CC inflammation. The present sequence is a human heparanase peptide used  
 XX to raise anti-heparanase antibodies.  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 82; DB 8; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGETSSAY 15  
 DB 1 RPKKKVWLGETSSAY 15  
 RESULT 2  
 ADR88215  
 ID ADR88215 standard; peptide; 15 AA.  
 XX  
 AC ADR88215;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE Human heparanase epitope pep9.  
 XX  
 KW Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;  
 KW autoimmune disorder; cancer; angiogenesis; metastatic disease;  
 KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;  
 KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;  
 KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;  
 KW human; heparanase; enzyme; epitope.  
 OS Homo sapiens.  
 XX  
 PN US2004170631-A1.  
 XX  
 PD 02-SEP-2004.  
 XX  
 XX 28-NOV-2003; 2003US-00722502.  
 PF  
 XX 02-SEP-1997; 97US-00922170.  
 PR 01-MAY-1998; 98US-00071739.  
 PR 04-NOV-1998; 98US-00186200.  
 PR 19-FEB-2003; 2003US-00368044.  
 PR 22-AUG-2003; 2003US-00645659.  
 XX  
 PA (YACO/) YACOBY-ZEEVI O.  
 PA (PERE/) PERETZ T.  
 PA (MIRO/) MIRON D.  
 PA (SHLO/) SHLOMI Y.

PA (PECK/) PECKER I.  
PA (AYAL/) AYAL-HERSHKOVITZ M.  
PA (FEIN/) FEINSTEIN E.  
PA (VSEL/) VAN GELDER J M.  
PA (VLOD/) VLODAVSKY I.  
PA (FRIE/) FRIEDMANN Y.  
XX  
PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
PI Ayal-Herskovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;  
PI Friedmann Y;  
XX  
DR WPI; 2004-625084/60.  
XX  
XX Targeted drug delivery to a heparanase-expressing tissue of a patient,  
PT useful for treating heparanase-associated conditions such as inflammation  
PT or cancer, comprises administering a drug and an anti-heparanase antibody  
PT complex.  
XX  
PS Claim 7; SEQ ID NO 9; 58pp; English.  
XX  
CC The invention relates to a method of targeted drug delivery to a tissue  
CC of a patient, the tissue expressing heparanase. The method comprises  
CC providing a complex of a drug directly or indirectly linked to an anti-  
CC heparanase antibody, and administering the complex to the patient. In the  
CC targeted drug delivery, the antibody comprises an antibody or its portion  
CC capable of specifically binding to at least one epitope of a heparanase  
CC protein. The composition and methods of the invention are useful for  
CC diagnosing, preventing or treating conditions associated with heparanase  
CC catalytic activity (e.g. an inflammatory disorder, wound, scar,  
CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell  
CC proliferation, invasion of circulating tumour cells and metastatic  
CC disease), for purifying heparanase, or for developing drugs for those  
CC heparanase-associated conditions. The vasculopathy is atherosclerosis,  
CC restenosis or aneurysm. The cancerous condition is a solid cancer or a  
CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy  
CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous  
CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous  
CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,  
CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and  
CC multiple myeloma. The solid cancer is selected from tumours in lip and  
CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,  
CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,  
CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of  
CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue  
CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,  
CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic  
CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary  
CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,  
CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,  
CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,  
CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's  
CC sarcoma. The present sequence is human heparanase epitope.  
XX  
SQ Sequence 15 AA;  
Query Match 100.0%; Score 82; DB 8; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.7e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 RPKKKVWLGETSSAY 15  
| | | | | | | | | | | | | | |  
Db 1 RPKKKVWLGETSSAY 15  
RESULT 3  
ADT78182  
ID ADT78182 standard; peptide; 15 AA.  
XX  
AC ADT78182;  
XX  
XX 13-JAN-2005 (first entry)  
XX  
DE Functional peptide epitope of human heparanase, pep9.

XX Antibody; epitope; heparanase; pathological condition; angiogenesis;  
KW cell proliferation; cancerous condition; tumour cell invasion;  
KW metastatic disease; heparanase-related disorder; inflammatory disorder;  
KW wound; scar; vasculopathy; autoimmune condition; renal disease;  
KW cytostatic; antiinflammatory; vulnery; antiarteriosclerotic;  
KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.  
XX  
OS Homo sapiens.  
XX  
PN US2004213789-A1.  
XX  
XX 28-OCT-2004.  
XX  
XX 22-AUG-2003; 2003US-00645659.  
PF  
XX 02-SEP-1997; 97US-00922170.  
PR 01-MAY-1998; 98US-00071739.  
PR 04-NOV-1998; 98US-00186200.  
PR 19-FEB-2003; 2003US-00368044.  
XX  
XX (YACO/) YACOBY-ZEEVI O.  
PA (PERE/) PERETZ T.  
PA (MIRO/) MIRON D.  
PA (SHLO/) SHLOMI Y.  
PA (PECK/) PECKER I.  
PA (AYAL/) AYAL-HERSHKOVITZ M.  
PA (FEIN/) FEINSTEIN E.  
PA (GELD/) GELDER J M V.  
PA (VLOD/) VLODAVSKY I.  
PA (FRIE/) FRIEDMANN Y.  
XX  
XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
PI Ayal-Herskovitz M, Feinstein E, Gelder JMW, Vlodavsky I;  
PI Friedmann Y;  
XX WPI; 2004-774790/76.  
DR  
XX New neutralizing monoclonal anti-heparanase antibodies, useful for  
PT detecting, treating or preventing cancer, inflammatory or autoimmune  
PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.  
XX  
PS Claim 67; SEQ ID NO 9; 68pp; English.  
XX  
CC The invention relates to an isolated antibody or antibody portion capable  
CC of specifically binding to or elicited by at least one epitope of a  
CC heparanase protein, where the heparanase protein is at least 60%  
CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and  
CC where at least one epitope comprises a sequence at least 70% homologous  
CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)  
CC a hybridoma cell line comprising a cell line for producing the monoclonal  
CC antibody, (b) a method for detecting, treating or preventing a  
CC pathological condition or a heparanase-related disorder or condition in a  
CC subject, (c) a method for monitoring the state of a heparanase-related  
CC disorder or condition in a subject, and (d) a pharmaceutical composition  
CC comprising the isolated anti-heparanase antibody or antibody portion and  
CC a pharmaceutical carrier. The antibody, methods, and composition are  
CC useful for detecting, treating, preventing or monitoring a pathological  
CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition  
CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,  
CC or prostate cancer), minor cell proliferation, invasion of circulating  
CC tumour cells, or a metastatic disease, or a heparanase-related disorder  
CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy  
CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or  
CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,  
CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell  
CC carcinoma) in a mammal. This sequence represents a functional peptide  
CC epitope of human heparanase.  
XX  
SQ Sequence 15 AA;  
Query Match 100.0%; Score 82; DB 8; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.7e-06;





XX Targeted drug delivery ; inflammatory disorder; wound; scar;  
 KW vasculopathy; autoimmune disorder; cancer; angiogenesis;  
 KW metastatic disease; atherosclerosis; restenosis; aneurysm; solid cancer;  
 KW non-solid cancer; haematopoietic malignancy ; lymphocytic leukaemia;  
 KW myelogenous leukaemia; Hodgkin's disease; multiple myeloma;  
 KW haemangiosarcoma; Kaposi's sarcoma; human ; heparanase; enzyme.  
 XX Homo sapiens.  
 OS  
 XX  
 XX US2004170631-A1.  
 XX  
 XX 02-SEP-2004.  
 XX  
 XX 28-NOV-2003; 2003US-00722502.  
 XX  
 XX 02-SEP-1997; 97US-00922170.  
 PR 01-MAY-1998; 98US-00071739.  
 PR 04-NOV-1998; 98US-00186200.  
 PR 19-FEB-2003; 2003US-00368044.  
 PR 22-AUG-2003; 2003US-00645659.  
 XX  
 XX (YACO/) YACOBY-ZEEVI O.  
 PA (PERE/) PERETZ T.  
 PA (MIRO/) MIRON D.  
 PA (SHLO/) SHLOMI Y.  
 PA (PECK/) PECKER I.  
 PA (AYAL/) AYAL-HERSHKOVITZ M.  
 PA (FEIN/) FEINSTEIN E.  
 PA (VGEL/) VAN GELDER J M.  
 PA (VLOD/) VLODAVSKY I.  
 PA (FRIE/) FRIEDMANN Y.  
 XX  
 XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
 PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;  
 PI Friedmann Y;  
 XX  
 XX WPI; 2004-625084/60.  
 XX  
 XX Targeted drug delivery to a heparanase-expressing tissue of a patient,  
 PT useful for treating heparanase-associated conditions such as inflammation  
 PT or cancer, comprises administering a drug and an anti-heparanase antibody  
 PT complex.  
 XX  
 XX Claim 2; SEQ ID NO 1; 58pp; English.  
 XX  
 XX The invention relates to a method of targeted drug delivery to a tissue  
 CC of a patient, the tissue expressing heparanase. The method comprises  
 CC providing a complex of a drug directly or indirectly linked to an anti-  
 CC heparanase antibody, and administering the complex to the patient. In the  
 CC targeted drug delivery, the antibody comprises an antibody or its portion  
 CC capable of specifically binding to at least one epitope of a heparanase  
 CC protein. The composition and methods of the invention are useful for  
 CC diagnosing, preventing or treating conditions associated with heparanase  
 CC catalytic activity (e.g. an inflammatory disorder, wound, scar,  
 CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell  
 CC proliferation, invasion of circulating tumour cells and metastatic  
 CC disease), for purifying heparanase, or for developing drugs for those  
 CC heparanase-associated conditions. The vasculopathy is atherosclerosis,  
 CC restenosis or aneurysm. The cancerous condition is a solid cancer or a  
 CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy  
 CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous  
 CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous  
 CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,  
 CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and  
 CC multiple myeloma. The solid cancer is selected from tumours in lip and  
 CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,  
 CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,  
 CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of  
 CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue  
 CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,  
 CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic  
 CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary

CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,  
 CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,  
 CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,  
 CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's  
 CC sarcoma. The present sequence is the 45 kDa major subunit of human mature  
 CC heparanase.  
 XX  
 XX Sequence 386 AA;  
 Query Match 100.0%; Score 82; DB 8; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 7.9e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RFGKKVWLGTSAY 15  
 DB 177 RFGKKVWLGTSAY 191  
 RESULT 7  
 ADT78174  
 ID ADT78174 standard; protein; 386 AA.  
 XX  
 XX ADT78174;  
 XX  
 XX 13-JAN-2005 (first entry)  
 XX  
 XX 45kDa subunit of mature processed human heparanase dimer.  
 KW Antibody; epitope; heparanase; pathological condition; angiogenesis;  
 KW cell proliferation; cancerous condition ; tumour cell invasion;  
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;  
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;  
 KW cytostatic; antiinflammatory; vulnery; antiarteriosclerotic;  
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.  
 XX  
 XX Homo sapiens.  
 XX  
 XX US2004213789-A1.  
 XX  
 XX 28-OCT-2004.  
 XX  
 XX 22-AUG-2003; 2003US-00645659.  
 XX  
 XX 02-SEP-1997; 97US-00922170.  
 PR 01-MAY-1998; 98US-00071739.  
 PR 04-NOV-1998; 98US-00186200.  
 PR 19-FEB-2003; 2003US-00368044.  
 XX  
 XX (YACO/) YACOBY-ZEEVI O.  
 PA (PERE/) PERETZ T.  
 PA (MIRO/) MIRON D.  
 PA (SHLO/) SHLOMI Y.  
 PA (PECK/) PECKER I.  
 PA (AYAL/) AYAL-HERSHKOVITZ M.  
 PA (FEIN/) FEINSTEIN E.  
 PA (VGEL/) GELDER J M V.  
 PA (VLOD/) VLODAVSKY I.  
 PA (FRIE/) FRIEDMANN Y.  
 XX  
 XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
 PI Ayal-Hershkovitz M, Feinstein E, Gelder JMV, Vlodavsky I;  
 PI Friedmann Y;  
 XX  
 XX WPI; 2004-774790/76.  
 XX  
 XX New neutralizing monoclonal anti-heparanase antibodies, useful for  
 PT detecting, treating or preventing cancer, inflammatory or autoimmune  
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.  
 XX  
 XX Claim 5; SEQ ID NO 1; 68pp; English.  
 PS  
 XX The invention relates to an isolated antibody or antibody portion capable  
 CC of specifically binding to or elicited by at least one epitope of a

CC heparanase protein, where the heparanase protein is at least 60%  
CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and  
CC where at least one epitope comprises a sequence at least 70% homologous  
CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)  
CC a hybridoma cell line comprising a cell line for producing the monoclonal  
CC antibody, (b) a method for detecting, treating or preventing a  
CC pathological condition or a heparanase-related disorder or condition in a  
CC subject, (c) a method for monitoring the state of a heparanase-related  
CC disorder or condition in a subject, and (d) a pharmaceutical composition  
CC comprising the isolated anti-heparanase antibody or antibody portion and  
CC a pharmaceutical carrier. The antibody, methods, and composition are  
CC useful for detecting, treating, preventing or monitoring a pathological  
CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition  
CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,  
CC or prostate cancer), minor cell proliferation, invasion of circulating  
CC tumour cells, or a metastatic disease, or a heparanase-related disorder  
CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy  
CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or  
CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,  
CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell  
CC carcinoma) in a mammal. This sequence represents the 45kDa subunit of  
CC mature processed human heparanase dimer.

XX  
SQ Sequence 386 AA;

Query Match 100.0%; Score 82; DB 8; Length 386;  
Best Local Similarity 100.0%; Pred. No. 7.9e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGETSSAY 15  
|||  
DB 177 RPKKKVWLGETSSAY 191

RESULT 8

ADY27057  
ID ADY27057 standard; protein; 386 AA.

AC  
XX ADY27057;

DT 05-MAY-2005 (first entry)

XX Heparanase inhibitor protein #1.

DE Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
KW neurological disease; viral infection; infection; cytostatic;  
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
KW heparanase modulator; enzyme purification.

XX Homo sapiens.

XX WO2005016227-A2.

XX 24-FEB-2005.

XX 12-AUG-2004; 2004WO-IL000744.

XX 14-AUG-2003; 2003US-0494800P.

XX 12-JAN-2004; 2004US-0535492P.

XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX Van-Gelder JM, Miron D;

XX WPI; 2005-182203/19.

XX Regulating heparanase activity, useful for treating heparanase-associated  
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
PT neurological diseases or viral diseases) comprises modulating heparanase  
PT activation.

XX Claim 55; SEQ ID NO 33; 211pp; English.

XX

CC The invention relates to a method of regulating heparanase activity in a  
CC tissue or regulating a biological process depending at least in part on  
CC heparanase activity comprising modulating heparanase activation. The  
CC invention also relates to methods of treating a heparanase- or heparin  
CC binding protein-associated disease or disorder in a subject, a  
CC pharmaceutical composition for use in the treatment of a heparanase-  
CC associated disease or disorder comprising a therapeutic amount of an  
CC agent capable of modulating heparanase activation and a pharmaceutical  
CC carrier or diluent, a method of identifying a protease activator of  
CC heparanase, a protease substrate mimetic comprising a peptide  
CC representing a subset or all substrate residues or cleavage sites of  
CC human heparanase or an equivalent non-human heparanase, a method of  
CC producing active heparanase and a method of modulating an adhesion  
CC activity of heparanase. The composition and methods are useful for  
CC modulating heparanase activation and for treating heparanase-associated  
CC diseases or disorders such as cancer, inflammation, cardiovascular  
CC diseases, neurological diseases or viral infections. This sequence  
CC represents a heparanase inhibitor protein used in the scope of the  
CC invention.

XX  
SQ Sequence 386 AA;

Query Match 100.0%; Score 82; DB 9; Length 386;  
Best Local Similarity 100.0%; Pred. No. 7.9e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGETSSAY 15  
|||  
DB 177 RPKKKVWLGETSSAY 191

RESULT 9

ADZ18995  
ID ADZ18995 standard; protein; 386 AA.

XX  
AC ADZ18995;

DT 16-JUN-2005 (first entry)

XX Human heparanase consensus cleavage site #2.

XX Enzyme engineering; heparanase; metastasis; autoimmune disease;  
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
KW immunosuppressive; enzyme.

XX Homo sapiens.

XX WO2005030962-A1.

XX 07-APR-2005.

XX 17-SEP-2004; 2004WO-EP010517.

XX 26-SEP-2003; 2003US-0506479P.

XX 20-JAN-2004; 2004US-0537729P.

XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.

XX Lahm A, Nardella C, Pallaro M, Steinkuhler C;

XX WPI; 2005-273382/28.

XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
PT nucleotide sequence that encodes mammalian heparanase protein and has two  
PT consensus cleavage sites located between specific nucleotide encoding  
PT residues.

XX Disclosure; SEQ ID NO 16; 65pp; English.

XX The invention relates to a synthetic nucleic acid molecule that encodes  
CC mammalian heparanase protein, where the nucleic acid comprises two  
CC consensus cleavage sites recognized by endoproteinase. The sequences are  
CC useful for expressing mammalian heparanase in non-mammalian cells and in

CC inhibitor screening assays for the development of therapeutics or  
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a human heparanase  
 CC consensus cleavage site used in the scope of the invention.

XX SQ Sequence 386 AA;

Query Match 100.0%; Score 82; DB 9; Length 386;

Best Local Similarity 100.0%; Pred. No. 7.9e-05; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0;

QY 1 RPCKVWLGETSSAY 15

|||||

Db 177 RPCKVWLGETSSAY 191

RESULT 10

AEA42423

ID AEA42423 standard; protein; 386 AA.

XX AC AEA42423;

XX DT 28-JUL-2005 (first entry)

XX DE Human mature heparanase dimer 45 kDa subunit SEQ ID NO:1.

XX KW antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;  
 KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;  
 KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;  
 KW angiogenesis disorder; cancer; tumor; metastasis.

XX OS Homo sapiens.

XX AU2004201462-A1.

XX PD 06-MAY-2004.

XX PF 08-APR-2004; 2004AU-00201462.

XX PR 08-APR-2004; 2004AU-00201462.

XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX PI Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;  
 PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Herschkovitz M, Ben-Artzi H;  
 PI Feinstein E;

XX WPI; 2005-173343/19.

XX Novel isolated antibody capable of specifically binding to epitope of  
 PT heparanase protein, useful for preventing and treating heparanase-related  
 PT disorder such as inflammatory disorder, scars, autoimmune conditions or  
 PT angiogenesis.

XX Claim 2; SEQ ID NO 1; 260pp; English.

XX The invention relates to an isolated antibody or its portion (I) capable  
 CC of specifically binding to an epitope of a heparanase protein. Also  
 CC described: (1) a cell line (II) for producing a monoclonal antibody or  
 CC its portion, comprising a cell line for producing (I); (2) a  
 CC pharmaceutical composition comprising (I) and a carrier; and (3) an  
 CC affinity medium (III) for binding human heparanase polypeptides.  
 CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)  
 CC useful for treating a subject suffering from a pathological condition,  
 CC which involves administering (I) to the subject. (I) is useful for  
 CC preventing and treating heparanase-related disorder or condition chosen  
 CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune  
 CC condition, angiogenesis, cell proliferation, cancerous condition, tumor  
 CC cell proliferation, invasion of circulating tumor cells and metastatic  
 CC disease. (I) is useful for detecting the presence of heparanase  
 CC polypeptide in a sample. (I) is useful for detecting heparanase-related  
 CC disease or condition in a subject such as vertebrate, preferably mammal

CC e.g., human. The heparanase-related disorder or condition further  
 CC includes renal disease or disorder chosen from diabetic nephropathy,  
 CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome  
 CC and renal cell carcinoma. The present sequence represents the 45 kDa  
 CC subunit of the human mature processed heparanase dimer, which is used in  
 CC the exemplification of the present invention.

XX SQ Sequence 386 AA;

Query Match 100.0%; Score 82; DB 9; Length 386;

Best Local Similarity 100.0%; Pred. No. 7.9e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPCKVWLGETSSAY 15

|||||

Db 177 RPCKVWLGETSSAY 191

RESULT 11

ADY27061

ID ADY27061 standard; protein; 460 AA.

XX AC ADY27061;

XX DT 05-MAY-2005 (first entry)

XX DE Heparanase inhibitor protein #4.

XX KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
 KW neurological disease; viral infection; infection; cytostatic;  
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
 KW heparanase modulator; enzyme purification.

XX OS Homo sapiens.

XX WO2005016227-A2.

XX PD 24-FEB-2005.

XX PF 12-AUG-2004; 2004WO-IL000744.

XX PR 14-AUG-2003; 2003US-0494800P.

XX PR 12-JAN-2004; 2004US-0535492P.

XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX Van-Gelder JM, Miron D;

XX WPI; 2005-182203/19.

XX Regulating heparanase activity, useful for treating heparanase-associated  
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
 PT neurological diseases or viral diseases) comprises modulating heparanase  
 PT activation.

XX Disclosure; SEQ ID NO 37; 211pp; English.

XX The invention relates to a method of regulating heparanase activity in a  
 CC tissue or regulating a biological process depending at least in part on  
 CC heparanase activity comprising modulating heparanase activation. The  
 CC invention also relates to methods of treating a heparanase- or heparin  
 CC binding protein-associated disease or disorder in a subject, a  
 CC pharmaceutical composition for use in the treatment of a heparanase-  
 CC associated disease or disorder comprising a therapeutic amount of an  
 CC agent capable of modulating heparanase activation and a pharmaceutical  
 CC carrier or diluent, a method of identifying a protease activator of  
 CC heparanase, a protease substrate mimetic comprising a peptide  
 CC representing a subset or all substrate residues or cleavage sites of  
 CC human heparanase or an equivalent non-human heparanase, a method of  
 CC producing active heparanase and a method of modulating an adhesion  
 CC activity of heparanase. The composition and methods are useful for  
 CC modulating heparanase activation and for treating heparanase-associated  
 CC diseases or disorders such as cancer, inflammation, cardiovascular

CC diseases, neurological diseases or viral infections. This sequence  
 CC represents a heparanase inhibitor protein used in the scope of the  
 CC invention.

XX SQ Sequence 460 AA;

Query Match 100.0%; Score 82; DB 9; Length 460;  
 Best Local Similarity 100.0%; Pred. No. 9.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGETSSAY 15  
 |||||  
 DB 251 RPKKKVWLGETSSAY 265

# RESULT 12

AE887589  
 ID AEB87589 standard; protein; 486 AA.

XX AC AEB87589;

XX DT 06-OCT-2005 (first entry)

XX DE Human heparanase 65delta20 deletion mutant.

XX KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;  
 KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;  
 KW leukemia; lymphoma; dermatological disease; hematological disease;  
 KW immune disorder; inflammation; antiinflammatory; renal disease;  
 KW nephrotropic; endocrine disease; genitourinary disease;  
 KW autoimmune disease; immunosuppressive; drug screening; mutagen.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO2005071070-A2.

XX PD 04-AUG-2005.

XX PF 20-JAN-2005; 2005WO-IL000068.

XX PR 22-JAN-2004; 2004IL-00160025.

XX PR 28-JUL-2004; 2004US-00901943.

XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX PI Vlodavsky I, Ilan N, Levy-Adam F;

XX DR WPI: 2005-564219/57.

XX DR N-PSDB; AEB87588.

XX PT New amino acid sequences derived from the 50 kDa subunit of heparanase,  
 PT for treating or inhibiting malignant proliferative, inflammatory, kidney  
 PT disorder or autoimmune disorder.

XX PS Claim 107; SEQ ID NO 31; 167pp; English.

XX CC The present sequence is that of a deletion mutant of human heparanase,  
 CC denoted 65delta20, which is devoid of amino acid residues 411-432 of the  
 CC native protein. The recombinant protein is deficient of heparanase  
 CC endoglycosidase catalytic activity. The invention relates to amino acid  
 CC sequences derived from the N-terminus region of the 50 kDa subunit of  
 CC heparanase, particularly in the regions between amino acid residues 158-  
 CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise  
 CC heparin-binding domains. The invention also provides an antibody directed  
 CC to these sequences, in particular the 158-171 peptide, and compositions  
 CC and uses of this antibody as a heparanase inhibitor. A screening method  
 CC is provided for specific heparanase inhibitors. Claimed pharmaceutical  
 CC compositions comprising (i) a peptide derived from the N-terminus region  
 CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a  
 CC peptide, or (iii) an antibody which specifically recognizes the peptide  
 CC are used for the inhibition of heparanase catalytic activity associated  
 CC with an inflammatory disorder, kidney disease, autoimmune disease,

CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or  
 CC with a malignant proliferative disorder, especially a solid or non-solid  
 CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.

XX SQ Sequence 486 AA;

Query Match 100.0%; Score 82; DB 9; Length 486;  
 Best Local Similarity 100.0%; Pred. No. 0.0001;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGETSSAY 15  
 |||||  
 DB 299 RPKKKVWLGETSSAY 313

# RESULT 13

ADZ18996  
 ID ADZ18996 standard; protein; 492 AA.

XX AC ADZ18996;

XX DT 16-JUN-2005 (first entry)

XX DE Hep106 construct protein.

XX KW Enzyme engineering; heparanase; hep106; metastasis; autoimmune disease;  
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
 KW immunosuppressive; enzyme.

XX OS Synthetic.

XX PN WO2005030962-A1.

XX PD 07-APR-2005.

XX PF 17-SEP-2004; 2004WO-EP010517.

XX PR 26-SEP-2003; 2003US-0506479P.

XX PR 20-JAN-2004; 2004US-0537729P.

XX PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.

XX PI Lahm A, Nardella C, Fallaoro M, Steinkuhler C;

XX DR WPI: 2005-273382/28.

XX DR N-PSDB; ADZ18997.

XX PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.

XX PS Example 2; SEQ ID NO 17; 65pp; English.

XX CC The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoproteinase. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in  
 CC inhibitor screening assays for the development of therapeutic or  
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a hep106 construct protein  
 CC used in the scope of the invention.

XX SQ Sequence 492 AA;

Query Match 100.0%; Score 82; DB 9; Length 492;  
 Best Local Similarity 100.0%; Pred. No. 0.0001;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGETSSAY 15  
 |||||  
 DB 283 RPKKKVWLGETSSAY 297

RESULT 14  
AEB87562  
ID AEB87562 standard; protein; 493 AA.  
XX AC AEB87562;  
XX DT 06-OCT-2005 (first entry)  
XX DE Human heparanase 65delta15 deletion mutant.  
XX KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;  
KW cystostatic; metastasis; hyperproliferation; carcinoma; sarcoma;  
KW leukemia; lymphoma; dermatological disease; hematological disease;  
KW immune disorder; inflammation; antiinflammatory; renal disease;  
KW nephrotropic; endocrine disease; genitourinary disease;  
KW autoimmune disease; immunosuppressive; drug screening; muten.  
XX OS Homo sapiens.  
OS Synthetic.  
XX PN WO2005071070-A2.  
XX PD 04-AUG-2005.  
XX PF 20-JAN-2005; 2005WO-IL000068.  
XX PR 22-JAN-2004; 2004IL-00160025.  
XX PR 28-JUL-2004; 2004US-00901943.  
XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.  
XX PI Vlodavsky I, Ilan N, Levy-Adam F;  
XX WPI; 2005-564219/57.  
XX DR N-PSDB; AEB87561.  
XX PT New amino acid sequences derived from the 50 kDa subunit of heparanase,  
PT for treating or inhibiting malignant proliferative, inflammatory, kidney  
PT disorder or autoimmune disorder.  
XX PS Claim 105; SEQ ID NO 4; 167pp; English.  
XX CC The present sequence is that of a deletion mutant of human heparanase,  
CC denoted 65delta15, which is devoid of amino acid residues 158-171 of the  
CC native protein. The recombinant protein is deficient of heparanase  
CC endoglycosidase catalytic activity. The invention relates to amino acid  
CC sequences derived from the N-terminus region of the 50 kDa subunit of  
CC heparanase, particularly in the regions between amino acid residues 158-  
CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise  
CC heparin-binding domains. The invention also provides an antibody directed  
CC to these sequences, in particular the 158-171 peptide, and compositions  
CC and uses of this antibody as a heparanase inhibitor. A screening method  
CC is provided for specific heparanase inhibitors. Claimed pharmaceutical  
CC compositions comprising (i) a peptide derived from the N-terminus region  
CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a  
CC peptide, or (iii) an antibody which specifically recognizes the peptide  
CC are used for the inhibition of heparanase catalytic activity associated  
CC with an inflammatory disorder, kidney disease, autoimmune disease,  
CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or  
CC with a malignant proliferative disorder, especially a solid or non-solid  
CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.  
XX SQ Sequence 493 AA;  
Query Match 100.0%; Score 82; DB 9; Length 493;  
Best Local Similarity 100.0%; Pred. No. 0.0001;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 RPKKKVWLGETSSAY 15  
Db 284 RPKKKVWLGETSSAY 298

RESULT 15  
ADZ18999  
ID ADZ18999 standard; protein; 495 AA.  
XX AC ADZ18999;  
XX DT 16-JUN-2005 (first entry)  
XX DE Hep109 construct protein.  
XX KW Enzyme engineering; heparanase; hep109; metastasis; autoimmune disease;  
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
KW immunosuppressive; enzyme.  
XX OS Synthetic.  
XX PN WO2005030962-A1.  
XX PD 07-APR-2005.  
XX PF 17-SEP-2004; 2004WO-EP010517.  
XX PR 26-SEP-2003; 2003US-0506479P.  
XX PR 20-JAN-2004; 2004US-0537729P.  
XX PA (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.  
XX PI Lahm A, Nardella C, Pallao M, Steinkuhler C;  
XX WPI; 2005-273382/28.  
XX DR N-PSDB; ADZ18998.  
XX PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
PT nucleotide sequence that encodes mammalian heparanase protein and has two  
PT consensus cleavage sites located between specific nucleotide encoding  
PT residues.  
XX PS Example 2; SEQ ID NO 20; 65pp; English.  
XX CC The invention relates to a synthetic nucleic acid molecule that encodes  
CC mammalian heparanase protein, where the nucleic acid comprises two  
CC consensus cleavage sites recognized by endoproteinase. The sequences are  
CC useful for expressing mammalian heparanase in non-mammalian cells and in  
CC inhibitor screening assays for the development of therapeutics or  
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
CC and/or inflammation. This sequence represents a hep109 construct protein  
CC used in the scope of the invention.  
XX SQ Sequence 495 AA;  
Query Match 100.0%; Score 82; DB 9; Length 495;  
Best Local Similarity 100.0%; Pred. No. 0.0001;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 RPKKKVWLGETSSAY 15  
Db 286 RPKKKVWLGETSSAY 300  
RESULT 16  
AEB87587  
ID AEB87587 standard; protein; 497 AA.  
XX AC AEB87587;  
XX DT 06-OCT-2005 (first entry)  
XX DE Human heparanase 65delta10 deletion mutant.  
XX KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;  
KW cystostatic; metastasis; hyperproliferation; carcinoma; sarcoma;  
KW leukemia; lymphoma; dermatological disease; hematological disease;  
KW

```

KW immune disorder; inflammation; antiinflammatory; renal disease;
KW nephrotropic; endocrine disease; genitourinary disease;
KW autoimmune disease; immunosuppressive; drug screening; mutin.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2005071070-A2.
XX
XX 04-AUG-2005.
XX
XX 20-JAN-2005; 2005WO-IL000068.
XX
XX 22-JAN-2004; 2004IL-00160025.
XX
XX 28-JUL-2004; 2004US-00901943.
XX
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
XX
XX Vlodavsky I, Ilan N, Levy-Adam F;
XX
XX WPI; 2005-564219/57.
XX
XX N-PSDB; AEB87586.
XX
XX New amino acid sequences derived from the 50 kDa subunit of heparanase,
XX for treating or inhibiting malignant proliferative, inflammatory, kidney
XX disorder or autoimmune disorder.
XX
XX Claim 106; SEQ ID NO 29; 167pp; English.
XX
XX The present sequence is that of a deletion mutant of human heparanase,
XX denoted 65delta10, which is devoid of amino acid residues 270-280 of the
XX native protein. The recombinant protein is deficient of heparanase
XX endoglycosidase catalytic activity. The invention relates to amino acid
XX sequences derived from the N-terminus region of the 50 kDa subunit of
XX heparanase, particularly in the regions between amino acid residues 158-
XX 171, 270-280 and 411-432 of human heparanase. These sequences comprise
XX heparin-binding domains. The invention also provides an antibody directed
XX to these sequences, in particular the 158-171 peptide, and compositions
XX and uses of this antibody as a heparanase inhibitor. A screening method
XX is provided for specific heparanase inhibitors. Claimed pharmaceutical
XX compositions comprising (i) a peptide derived from the N-terminus region
XX of the heparanase 50 kDa subunit, (ii) a substance which binds to such a
XX peptide, or (iii) an antibody which specifically recognizes the peptide
XX are used for the inhibition of heparanase catalytic activity associated
XX with an inflammatory disorder, kidney disease, autoimmune disease,
XX angiogenesis, tumor formation, tumor progression or tumor metastasis, or
XX with a malignant proliferative disorder, especially a solid or non-solid
XX tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.
XX
XX Sequence 497 AA;

Query Match          100.0%; Score 82; DB 9; Length 497;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  RPKKKVWLGETSSAY 15
Db      288 RPKKKVWLGETSSAY 302
      |||||
RESULT 17
ADZ19000
ID      ADZ19000 standard; protein; 501 AA.
XX
XX ADZ19000;
XX
XX 16-JUN-2005 (first entry)
XX
XX HepGS3 construct protein.
XX
XX Enzyme engineering; heparanase; hepGS3; metastasis; autoimmune disease;
XX inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
XX immunosuppressive; enzyme.
XX
XX Synthetic.
XX
XX WO2005030962-A1.
XX
XX 07-APR-2005.
XX
XX 17-SEP-2004; 2004WO-EP010517.
XX
XX 26-SEP-2003; 2003US-0506479P.
XX
XX 20-JAN-2004; 2004US-0537729P.
XX
XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
XX Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
XX
XX WPI; 2005-273382/28.
XX
XX N-PSDB; ADZ19001.
XX
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
XX nucleotide sequence that encodes mammalian heparanase protein and has two
XX consensus cleavage sites located between specific nucleotide encoding
XX residues.
XX
XX Example 2; SEQ ID NO 21; 65pp; English.
XX
XX The invention relates to a synthetic nucleic acid molecule that encodes
XX mammalian heparanase protein, where the nucleic acid comprises two
XX consensus cleavage sites recognized by endoproteinase. The sequences are
XX useful for expressing mammalian heparanase in non-mammalian cells and in
XX inhibitor screening assays for the development of therapeutics or
XX pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
XX and/or inflammation. This sequence represents a hepGS3 construct protein
XX used in the scope of the invention.
XX
XX Sequence 501 AA;

Query Match          100.0%; Score 82; DB 9; Length 501;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  RPKKKVWLGETSSAY 15
Db      292 RPKKKVWLGETSSAY 306
      |||||
RESULT 18
ADZ19005
ID      ADZ19005 standard; protein; 507 AA.
XX
XX ADZ19005;
XX
XX 16-JUN-2005 (first entry)
XX
XX HepGS6 construct protein.
XX
XX Enzyme engineering; heparanase; hepGS6; metastasis; autoimmune disease;
XX inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
XX immunosuppressive; enzyme.
XX
XX Synthetic.
XX
XX WO2005030962-A1.
XX
XX 07-APR-2005.
XX
XX 17-SEP-2004; 2004WO-EP010517.
XX
XX 26-SEP-2003; 2003US-0506479P.
XX
XX 20-JAN-2004; 2004US-0537729P.
XX
XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
XX

```



PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;  
 XX WPI; 2005-273382/28.  
 DR N-PSDB; ADZ19003.  
 XX  
 PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.  
 XX  
 PS Example 2; SEQ ID NO 26; 65pp; English.  
 XX  
 CC The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoproteinase. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in  
 CC inhibitor screening assays for the development of therapeutics or  
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a hepgS6 construct protein  
 CC used in the scope of the invention.  
 XX  
 SQ Sequence 507 AA;  
 Query Match 100.0%; Score 82; DB 9; Length 507;  
 Best Local Similarity 100.0%; Pred. No. 0.00011;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPKKKVWLGETSSAY 15  
 DB 298 RPKKKVWLGETSSAY 312  
 |||||  
 RESULT 19  
 ADY27058  
 ID ADY27058 standard; protein; 508 AA.  
 XX  
 AC ADY27058;  
 XX  
 DT 05-MAY-2005 (first entry)  
 XX  
 DE Human inactive heparanase protein.  
 XX  
 KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
 KW neurological disease; viral infection; infection; cytostatic;  
 KW antinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
 KW protease; enzyme; enzyme purification.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2005016227-A2.  
 XX  
 PD 24-FEB-2005.  
 XX  
 PF 12-AUG-2004; 2004WO-IL000744.  
 XX  
 PR 14-AUG-2003; 2003US-0494800P.  
 PR 12-JAN-2004; 2004US-0535492P.  
 XX  
 PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.  
 XX  
 PI Van-Gelder JM, Miron D;  
 XX  
 DR WPI; 2005-182203/19.  
 XX  
 CC Regulating heparanase activity, useful for treating heparanase-associated  
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
 PT neurological diseases or viral diseases) comprises modulating heparanase  
 PT activation.  
 XX  
 PS Claim 257; SEQ ID NO 34; 21lpp; English.  
 XX  
 CC The invention relates to a method of regulating heparanase activity in a  
 CC tissue or regulating a biological process depending at least in part on

CC heparanase activity comprising modulating heparanase activation. The  
 CC invention also relates to methods of treating a heparanase- or heparin  
 CC binding protein-associated disease or disorder in a subject, a  
 CC pharmaceutical composition for use in the treatment of a heparanase-  
 CC associated disease or disorder comprising a therapeutic amount of an  
 CC agent capable of modulating heparanase activation and a pharmaceutical  
 CC carrier or diluent, a method of identifying a protease activator of  
 CC heparanase, a protease substrate mimetic comprising a peptide  
 CC representing a subset or all substrate residues or cleavage sites of  
 CC human heparanase or an equivalent non-human heparanase, a method of  
 CC producing active heparanase and a method of modulating an adhesion  
 CC activity of heparanase. The composition and methods are useful for  
 CC modulating heparanase activation and for treating heparanase-associated  
 CC diseases or disorders such as cancer, inflammation, cardiovascular  
 CC diseases, neurological diseases or viral infections. This sequence  
 CC represents a human inactive heparanase protein used in the scope of the  
 CC invention.  
 XX  
 SQ Sequence 508 AA;  
 Query Match 100.0%; Score 82; DB 9; Length 508;  
 Best Local Similarity 100.0%; Pred. No. 0.00011;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPKKKVWLGETSSAY 15  
 DB 299 RPKKKVWLGETSSAY 313  
 |||||  
 RESULT 20  
 ADZ19006  
 ID ADZ19006 standard; protein; 526 AA.  
 XX  
 AC ADZ19006;  
 XX  
 DT 16-JUN-2005 (first entry)  
 XX  
 DE HepHyaluro construct protein.  
 XX  
 KW Enzyme engineering; heparanase; hepHyaluro; metastasis;  
 KW autoimmune disease; inflammation; neoplasm; immune disorder;  
 KW antinflammatory; cytostatic; immunosuppressive; enzyme.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2005030962-A1.  
 XX  
 PD 07-APR-2005.  
 XX  
 PF 17-SEP-2004; 2004WO-EP010517.  
 XX  
 PR 26-SEP-2003; 2003US-0506479P.  
 PR 20-JAN-2004; 2004US-0537729P.  
 XX  
 PA (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.  
 XX  
 PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;  
 XX  
 DR WPI; 2005-273382/28.  
 DR N-PSDB; ADZ19007.  
 XX  
 XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.  
 XX  
 PS Example 2; SEQ ID NO 27; 65pp; English.  
 XX  
 CC The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoproteinase. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in  
 CC inhibitor screening assays for the development of therapeutics or

CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
CC and/or inflammation. This sequence represents a hephyaluro construct  
CC protein used in the scope of the invention.

XX SQ Sequence 526 AA;

Query Match 100.0%; Score 82; DB 9; Length 526;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGETSSAY 15  
| | | | | | | | | | | | | | | | | |  
Db 317 RPKKKVWLGETSSAY 331

RESULT 21

ABB07815 ABB07815 standard; protein; 527 AA.

XX AC ABB07815;

XX DT 03-JUL-2002 (first entry)

XX DE Chicken signal peptide/human heparanase chimeric protein sequence.

XX Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;  
KW anti-protozoan; neuroprotective; heparin; chicken; human; chimeric.

XX Synthetic.

OS Gallus gallus.

OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide

FT /note= "chicken heparanase signal peptide"

FT Protein

FT /note= "human heparanase mature protein"

XX US2002034810-A1.

XX PD 21-MAR-2002.

XX PF 16-AUG-2001; 2001US-00930218.

XX PR 20-SEP-2000; 2000US-00666390.

XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX PI Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;

XX DR WPI; 2002-338926/37.

XX DR N-PSDB; ABL40753.

XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful  
PT to treat various heparin-related disorders and the signal peptide is  
PT useful in production of membrane-targeted or secreted recombinant  
PT proteins.

XX PS Disclosure; Page 26-28; 39pp; English.

XX The invention relates to an isolated avian and reptile nucleic acid,  
CC encoding a polypeptide with heparanase catalytic activity. The signal  
CC peptide of the nucleic acid can be used to express membrane-associated or  
CC secreted proteins in heterologous expression systems. The encoded  
CC polypeptides can be used to prevent tumour angiogenesis, metastasis and  
CC invasion, and to intervene with pathologies associated with impaired  
CC heparin-binding growth factors, cellular responses to heparin-binding  
CC growth factors and cytokines, cell interaction with plasma lipoproteins,  
CC cellular susceptibility to viral, protozoa and bacterial infections or  
CC disintegration of neurodegenerative plaques. The present sequence  
CC represents a chicken signal peptide/human heparanase chimeric protein  
CC sequence

SQ Sequence 527 AA;

Query Match 100.0%; Score 82; DB 5; Length 527;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGETSSAY 15  
| | | | | | | | | | | | | | | | | |  
Db 318 RPKKKVWLGETSSAY 332

RESULT 22

ABW02018 ABW02018 standard; protein; 527 AA.

XX AC ABW02018;

XX DT 12-FEB-2004 (first entry)

XX DE Chimeric human-chicken heparanase protein.

XX KW Chicken; heparanase; tumour cell metastasis; inflammation; autoimmunity;  
KW wound healing; angiogenesis; restenosis; Genstmann-Straussler Syndrome;  
KW neurodegenerative disease; atherosclerosis; Creutzfeldt-Jakob disease;  
KW infection; Scrapie; Alzheimer's disease; protein therapy; cytostatic;  
KW immunosuppressive; vulnery; bactericide; anti-angiogenic; virucide;  
KW antisclerotic; neuroprotective; protozoacide; chimeric; fusion protein;  
KW enzyme; human.

XX OS Chimeric - Gallus gallus.

XX OS Chimeric - Homo sapiens.

XX PN US2003180788-A1.

XX PD 25-SEP-2003.

XX PF 08-MAY-2003; 2003US-00431438.

XX PR 20-SEP-2000; 2000US-00666390.

XX PR 16-AUG-2001; 2001US-00930218.

XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX PI Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;

XX DR WPI; 2003-843931/78.

XX DR N-PSDB; AAD63532.

XX Recombinant jungle red fowl (Gallus gallus) heparanase protein, useful  
PT for treating cancers, microbial infections and aiding wound healing.

XX PS Example; Page 26-28; Opp; English.

XX The present invention relates to novel jungle red fowl heparanase protein  
CC and polynucleotides encoding such proteins. Heparanase sequences can be  
CC used to develop treatments for various diseases, to develop diagnostic  
CC assays for these diseases and to provide new tools for basic and directed  
CC research especially in the fields of medicine and biology. They can be  
CC used to develop new drugs to inhibit tumour cell metastasis, inflammation  
CC and autoimmunity. Recombinant heparanase offers a potential treatment for  
CC wound healing, angiogenesis, restenosis, atherosclerosis, inflammation,  
CC neurodegenerative diseases (e.g. Genstmann-Straussler Syndrome, Scrapie,  
CC Creutzfeldt-Jakob disease and Alzheimer's disease) and certain viral and  
CC some bacterial and protozoa infections. Recombinant heparanase can also  
CC be used to neutralise plasma heparin, as a potential replacement of  
CC protamine. Sequences of the invention are used in protein therapy. The  
CC present sequence is chimeric human-chicken heparanase protein

XX SQ Sequence 527 AA;

Query Match 100.0%; Score 82; DB 7; Length 527;  
Best Local Similarity 100.0%; Pred. No. 0.00011;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPQKKVWLGETSSAY 15  
 |||||  
 Db 318 RPQKKVWLGETSSAY 332

RESULT 23  
 ADO63825 standard; protein; 527 AA.  
 XX ADO63825;  
 XX 26-AUG-2004 (first entry)  
 XX Chimeric heparanase mutant E225A, SEQ ID:10.  
 XX Human; chicken; heparanase; heparanase-derived protein;  
 KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;  
 KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;  
 KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnery; mutant; mutein.  
 XX Homo sapiens.  
 OS Gallus gallus.  
 OS Synthetic.  
 OS Chimeric.

Key Location/Qualifiers  
 FT Peptide 1..18 /note= "Chicken heparanase signal peptide"  
 FT Region 19..527  
 FT /note= "Corresponds to residues 35-543 of human  
 FT heparanase mutant E225A (SEQ ID NO:7)"  
 FT Misc-difference 209  
 FT /note= "Ala replaces wild-type Glu (active site proton  
 FT donor). Corresponds to residue 225 of human heparanase  
 FT mutant E225A (SEQ ID NO:7)"  
 FT Active-site 327  
 FT /note= "Active site nucleophile"  
 XX W02004048558-A2.  
 XX 10-JUN-2004.  
 XX 24-NOV-2003; 2003WO-IL000989.  
 XX 24-NOV-2002; 2002IL-00153059.  
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
 XX WPI; 2004-450373/42.  
 XX N-PSDB; ADO63819.  
 XX New nucleic acid construct comprising heparanase-derived polypeptide,  
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 PT thrombasthenia, or Bernard-Soulier syndrome.  
 XX Claim 10; SEQ ID NO 10; 128pp; English.  
 XX The invention relates to nucleic acid constructs comprising a nucleic  
 CC acid encoding a heparanase-derived protein which lacks heparanase  
 CC endoglycosidase catalytic activity but which retains its cell-cell and  
 CC cell-matrix adhesion properties. The constructs of the invention  
 CC optionally further comprise operably linked regulatory elements. The  
 CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the  
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343

are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 optionally further comprise an avian heparanase signal peptide (ADO63825-  
 ADO63827). The heparanase-derived protein, nucleic acid construct and  
 host cells are useful in preparing a tissue sealant composition for  
 sealing injuries, reducing the loss of blood, accelerating the healing  
 and homeostasis of an injury, accelerating blood vessel endothelium  
 formation or the endothelialisation of vascular grafts, accelerating the  
 adhesive activity of mammalian cells, and accelerating the adhesion and  
 aggregation of platelets. They may also be use in the treatment of  
 disorders associated with adhesion deficiency such as LAD (leukocyte  
 adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 cells of the invention may additionally be to screen for modulators of  
 cell-cell and cell-matrix adhesion, and to prepare an implantable  
 synthetic vascular graft comprising a tube made of a biocompatible  
 material lined with the cells. The present sequence represents a chimeric  
 protein comprising the signal peptide of chicken heparanase and residues  
 35-543 of the human heparanase mutant E225A.

Query Match 100.0%; Score 82; DB 8; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 0.00011;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPQKKVWLGETSSAY 15  
 |||||  
 Db 318 RPQKKVWLGETSSAY 332

RESULT 24  
 ADO63825 standard; protein; 527 AA.  
 XX ADO63825;  
 XX 16-JUN-2005 (first entry)  
 XX HepGS4 construct protein.  
 DE Enzyme engineering; heparanase; hepGS4; metastasis; autoimmune disease;  
 KW inflammation; neoplasm; immune disorder; antinflammatory; cytostatic;  
 KW immunosuppressive; enzyme.  
 XX Synthetic.  
 XX W02005030962-A1.  
 XX 07-APR-2005.  
 XX 17-SEP-2004; 2004WO-EP010517.  
 XX 26-SEP-2003; 2003US-0506479P.  
 XX 20-JAN-2004; 2004US-0537729P.  
 XX (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.  
 XX Lahm A, Nardella C, Pallaro M, Steinkuhler C;  
 XX WPI; 2005-273382/28.  
 XX N-PSDB; ADZ19002.  
 XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.  
 XX Example 2; SEQ ID NO 25; 65pp; English.  
 XX The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoproteinase. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in

CC inhibitor screening assays for the development of therapeutics or  
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a hepgs4 construct protein  
 CC used in the scope of the invention.

XX  
 SQ Sequence 527 AA;  
 Query Match 100.0%; Score 82; DB 9; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 0.00011;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPCKKVLGETSSAY 15  
 |||||  
 Db 318 RPCKKVLGETSSAY 332

RESULT 25  
 AAY34173  
 ID AAY34173 standard; protein; 530 AA.  
 XX  
 AC AAY34173;  
 XX  
 DT 15-NOV-1999 (first entry)  
 XX  
 DE Human pre-proheparanase protein sequence.  
 XX  
 KW Human; pre-proheparanase; platelet; wound healing; angiogenesis blocker;  
 KW inflammation; psoriasis; diabetic retinopathy; solid tumour; arthritis;  
 KW heparin degradation; anticoagulant neutralisation; asthma; CNS disease;  
 KW inflammatory disease; vascular restenosis; atherosclerosis; diagnosis;  
 KW tumour growth; fibroproliferative disorder; neurodegenerative disease;  
 KW therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9943830-A2.  
 XX  
 PD 02-SEP-1999.  
 XX  
 PF 18-FEB-1999; 99WO-US001489.  
 XX  
 PR 24-FEB-1998; 98US-0075706P.  
 PR 26-MAR-1998; 98US-0079401P.  
 XX  
 PA (PHAA ) PHARMACIA & UPJOHN CO.  
 XX  
 PI Heinrichson RL, Fairbanks MB, Mildner AM;  
 XX  
 DR WPI; 1999-540598/45.  
 DR N-PSDB; AAZ11236.  
 XX  
 FT New isolated platelet heparanase polypeptides, used to develop products  
 PT for, e.g. wound healing and blocking angiogenesis.  
 XX  
 PS Claim 12; Fig 7; 57pp; English.  
 XX  
 CC This sequence is the human pre-proheparanase of the invention. This  
 CC sequence was isolated from human platelets. The heparanase can be used  
 CC for identifying agents which alter heparanase activity. The heparanase  
 CC can be used for wound healing or for blocking angiogenesis or  
 CC inflammation. It can be used for treating e.g. psoriasis, diabetic  
 CC retinopathy or solid tumours, or for the degradation of heparin and the  
 CC neutralisation of heparin's anticoagulant properties during surgery.  
 CC Inhibitors of heparanase activity can be used in the treatment of  
 CC arthritis, asthma, and other inflammatory diseases, vascular restenosis,  
 CC atherosclerosis, tumour growth and progression, fibroproliferative  
 CC disorders, and central nervous system (CNS) and neurodegenerative  
 CC diseases. The products can also be used for detection and diagnosis. The  
 CC purified heparanase, both recombinantly produced human heparanase and  
 CC heparanase isolated from human platelet activity, allows for the  
 CC convenient selection of compounds having anti-heparanase activity, i.e.  
 CC inhibitors of heparanase activity, by measuring inhibition of heparanase  
 CC activity. Inhibition of heparanase activity can be measured by blocking

CC heparanase-mediated release of radioactive fragments from in vivo  
 CC radiolabelled (HSPG)/heparin  
 XX  
 SQ Sequence 530 AA;  
 Query Match 100.0%; Score 82; DB 2; Length 530;  
 Best Local Similarity 100.0%; Pred. No. 0.00011;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPCKKVLGETSSAY 15  
 |||||  
 Db 321 RPCKKVLGETSSAY 335

RESULT 26  
 AAY17083  
 ID AAY17083 standard; protein; 532 AA.  
 XX  
 AC AAY17083;  
 XX  
 DT 21-JUL-1999 (first entry)  
 XX  
 DE Seq ID No: 15 of WO9921975.  
 XX  
 KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;  
 KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;  
 KW arteriosclerosis; atherosclerosis; inflammation; tissue development;  
 KW human; HSPG.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9921975-A1.  
 XX  
 PD 06-MAY-1999.  
 XX  
 PF 28-OCT-1998; 98WO-AU000898.  
 XX  
 PR 28-OCT-1997; 97AU-00000062.  
 PR 09-DEC-1997; 97AU-00000812.  
 XX  
 PA (AUSU ) UNIV AUSTRALIAN NAT.  
 XX  
 PI Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;  
 XX  
 DR WPI; 1999-312956/26.  
 DR N-PSDB; AAX37260.  
 XX  
 PT Polynucleotides encoding mammalian endoglucuronidases, especially  
 PT heparanases, useful to promote wound healing.  
 XX  
 PS Claim 6; Page 76-79; 112pp; English.  
 XX  
 CC The invention relates to nucleic acid sequences that encode heparanase  
 CC enzymes having endoglucuronidase activity. Recombinant heparanases are  
 CC capable of removing the HS side chain from heparan sulfate proteoglycan  
 CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to  
 CC inhibit heparanase, this is useful for treatment of a physiological or  
 CC medical condition associated with elevated heparanase activity, such as  
 CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,  
 CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and  
 CC rat heparanases can be used to enhance wound healing, especially  
 CC associated with tissue development and repair. The conditions mentioned  
 CC above can be diagnosed using specific antibodies, and also using primers  
 CC and probes specific for the heparanase polynucleotides. Other uses of the  
 CC heparanases include sequencing sulfated molecules such as HSPG

XX  
 SQ Sequence 532 AA;  
 Query Match 100.0%; Score 82; DB 2; Length 532;  
 Best Local Similarity 100.0%; Pred. No. 0.00011;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPCKKVLGETSSAY 15

```

Db          334  RPCKKVLGETSSAY 348
|||||
RESULT 27
ID AAY02345
AA Y02345 standard; protein; 543 AA.
XX
XX
AC AAY02345;
XX
XX 09-JUL-1999 (first entry)
XX
XX A human heparanase protein.
XX
XX Heparanase; hp; modulator; heparin-binding growth factor;
KW cellular response; cytokine; cell interaction; plasma lipoprotein;
KW cellular susceptibility; infection; disintegration;
KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;
KW plasma heparin; micrometastasis; autoimmune lesion; renal failure.
XX
XX Homo sapiens.
OS
XX
XX WO9911798-A1.
PN
XX
XX 11-MAR-1999.
PD
XX
XX 31-AUG-1998; 98WO-US017954.
PF
XX
XX 02-SEP-1997; 97US-00922170.
PR
XX 02-JUL-1998; 98US-00109386.
PR
XX
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA
XX (FRIE/) FRIEDMAN M M.
PA
XX
XX Pecker I, Vlodavsky I, Feinstein E;
PI
XX
XX WPI; 1999-302255/25.
DR
XX N-PSDB; AAX35648.
DR
XX
XX New human polynucleotide useful for treating angiogenesis, restenosis,
PT and inflammation.
PT
XX
XX Claim 6; Fig 1; 63pp; English.
PS
XX
XX The specification describes a polypeptide having heparanase (hp)
CC activity. The recombinant protein is used as a modulator of heparin-
CC binding growth factors, cellular responses to heparin-binding growth
CC factors and cytokines, cell interaction with plasma lipoproteins,
CC cellular susceptibility to viral, protozoal and bacterial infections or
CC disintegration of neurodegenerative plaques. Heparanase may be useful for
CC conditions such as wound healing, angiogenesis, restenosis,
CC atherosclerosis, inflammation, neurodegenerative diseases, and viral
CC infections. Mammalian heparanase can be used to neutralize plasma
CC heparin, and anti-heparanase antibodies may be applied for
CC immunodetection and diagnosis of micrometastases, autoimmune lesions, and
CC renal failure in biopsy specimens, plasma samples, and body fluids. The
CC present sequence represents human heparanase
XX
XX Sequence 543 AA;
SQ
Query Match 100.0%; Score 82; DB 2; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPCKKVLGETSSAY 15
|||||
Db 334 RPCKKVLGETSSAY 348
|||||
RESULT 28
AAY17082

```

```

ID AAY17082 standard; protein; 543 AA.
XX
AC AAY17082;
XX
XX 21-JUL-1999 (first entry)
XX
XX Human heparanase enzyme.
XX
XX Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
KW arteriosclerosis; atherosclerosis; inflammation; tissue development;
KW human; HSPG.
XX
XX Homo sapiens.
OS
XX WO9921975-A1.
PN
XX
XX 06-MAY-1999.
PD
XX
XX 28-OCT-1998; 98WO-AU000898.
PF
XX
XX 28-OCT-1997; 97AU-00000062.
PR
XX 09-DEC-1997; 97AU-00000812.
PR
XX (AUSU ) UNIV AUSTRALIAN NAT.
PA
XX
XX Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;
PI
XX
XX WPI; 1999-312956/26.
DR
XX N-PSDB; AAX37259.
DR
XX
XX Polynucleotides encoding mammalian endoglucuronidases, especially
PT heparanases, useful to promote wound healing.
PT
XX
XX Claim 6; Page 69-73; 112pp; English.
PS
XX
XX The invention relates to nucleic acid sequences that encode heparanase
CC enzymes having endoglucuronidase activity. Recombinant heparanases are
CC capable of removing the HS side chain from heparan sulfate proteoglycan
CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
CC inhibit heparanase, this is useful for treatment of a physiological or
CC medical condition associated with elevated heparanase activity, such as
CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and
CC rat heparanases can be used to enhance wound healing, especially
CC associated with tissue development and repair. The conditions mentioned
CC above can be diagnosed using specific antibodies, and also using primers
CC and probes specific for the heparanase polynucleotides. Other uses of the
CC heparanases include sequencing sulfated molecules such as HSPG. The
CC present sequence represents a human heparanase
XX
XX Sequence 543 AA;
SQ
Query Match 100.0%; Score 82; DB 2; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPCKKVLGETSSAY 15
|||||
Db 334 RPCKKVLGETSSAY 348
|||||
RESULT 29
AAY57590
ID AAY57590 standard; protein; 543 AA.
XX
XX AAY57590;
AC
XX
XX 02-MAR-2000 (first entry)
DT
XX
XX Human heparanase.
DE
XX
XX Human; heparanase; hpa; genetic modification; expression; anticancer;
KW

```

KW angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumour;  
 KW anti-atherosclerotic; anti-inflammatory; antineurodegeneration;  
 KW heparan sulphate; heparin-binding growth factor; tumour angiogenesis;  
 KW metastasis; wound healing; restenosis; atherosclerosis; inflammation;  
 KW neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;  
 KW micrometastasis; autoimmune lesion; kidney failure.  
 OS Homo sapiens.  
 XX  
 XX WO9957244-A1.  
 XX  
 XX PD 11-NOV-1999.  
 XX  
 XX PF 29-APR-1999; 99WO-US0009256.  
 XX  
 XX PR 01-MAY-1998; 98US-00071618.  
 XX  
 XX PR 02-MAR-1999; 99US-00260038.  
 XX  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (FRIE/) FRIEDMAN M M.  
 XX  
 XX PA Ben-Artzi H, Ayal-Hershkovitz M, Yacoby-Zeevi O, Pecker I;  
 XX Peleg Y, Shlomi Y;  
 XX  
 XX WPI; 2000-062144/05.  
 DR N-PSDB; AA239195.  
 DR  
 XX Engineered cells that express recombinant heparanase, useful  
 PT therapeutically, e.g. for treating angiogenesis and to screen for  
 PT specific inhibitors, potential anticancer agents.  
 PT  
 XX Claim 3; Page 107-109; 118pp; English.  
 PS  
 XX The present invention describes genetically modified cells (A) containing  
 CC a polynucleotide (I) that encodes a polypeptide with heparanase activity,  
 CC and express recombinant heparanase (II). Heparanase cleaves heparan  
 CC sulphate (HS) at specific intrachain sites, resulting in release of  
 CC heparin-binding growth factors, enzymes and proteins that are sequestered  
 CC by HS in basement membranes, extracellular matrix or cell surfaces. It  
 CC may also be implicated in tumour angiogenesis and metastases. (II) is  
 CC potentially useful in wound healing and for treating angiogenesis,  
 CC restenosis, atherosclerosis, inflammation, neurodegeneration, viral  
 CC infection and cystic fibrosis. It can also be used to neutralise heparin  
 CC (an alternative to protamine) and to screen for specific inhibitors  
 CC (potentially useful for treating cancer and metastases). Antibodies  
 CC raised against (II) are used for immunodetection and diagnosis of  
 CC micrometastases, autoimmune lesions and kidney failure. (A) provide (II)  
 CC in large quantities, in a form that is homogeneously processed and  
 CC activated/neutralised by a dedicated protease. The present sequence  
 CC represents human heparanase  
 XX  
 XX Sequence 543 AA;  
 SQ  
 Query Match 100.0%; Score 82; DB 3; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 0.00011;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPKKKVWLGETSSAY 15  
 |||||  
 Db 334 RPKKKVWLGETSSAY 348  
 RESULT 30  
 AAB08849  
 ID AAB08849 standard; protein; 543 AA.  
 XX  
 XX AC AAB08849;  
 XX  
 XX DT 15-JAN-2001 (first entry)  
 XX  
 XX DE Amino acid sequence of a human heparanase polypeptide.  
 XX  
 XX KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;

KW heparin-binding growth factor; cytokine; neurodegenerative plaque;  
 KW wound healing; infection; burn; angiogenesis; restenosis;  
 KW atherosclerosis; inflammation; neurodegenerative disease;  
 KW Gerstmann-Strausser Syndrome; Creutzfeldt-Jakob disease.  
 OS Homo sapiens.  
 XX  
 XX PN WO200052178-A1.  
 XX  
 XX PD 08-SEP-2000.  
 XX  
 XX PF 14-FEB-2000; 2000WO-US003542.  
 XX  
 XX PR 01-MAR-1999; 99US-00258892.  
 XX  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA (FRIE/) FRIEDMAN M M.  
 XX  
 XX PI Pecker I, Vlodaysky I, Feinstein E;  
 XX  
 XX WPI; 2000-579289/54.  
 DR N-PSDB; AA75051.  
 DR  
 XX New polynucleotides encoding a polypeptide having heparanase activity,  
 PT useful in wound healing and in gene therapy, particularly in treating  
 PT tumor, inflammation, autoimmunity, neurodegenerative diseases.  
 PT  
 XX Claim 22; Fig 1; 152pp; English.  
 PS  
 XX The present sequence represents a human protein with heparanase catalytic  
 CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,  
 CC particularly in treating tumour, inflammation or autoimmunity.  
 CC Particularly, the polynucleotide is useful in modulating the  
 CC bioavailability of heparin-binding growth factors, cellular responses to  
 CC heparin-binding growth factors (e.g. bFGF) and cytokines (e.g.  
 CC interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular  
 CC susceptibility to certain viral and some bacterial and protozoa  
 CC infections, or disintegration of neurodegenerative plaques. The  
 CC polynucleotide is also useful in wound healing (e.g. thermal, chemical or  
 CC radiation burns), and in the treatment of angiogenesis, restenosis,  
 CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-  
 CC Strausler Syndrome or Creutzfeldt-Jakob disease), and some viral,  
 CC bacterial or protozoa infections  
 XX  
 XX Sequence 543 AA;  
 SQ  
 Query Match 100.0%; Score 82; DB 3; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 0.00011;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPKKKVWLGETSSAY 15  
 |||||  
 Db 334 RPKKKVWLGETSSAY 348  
 RESULT 31  
 AAY52990  
 ID AAY52990 standard; protein; 543 AA.  
 XX  
 XX AC AAY52990;  
 XX  
 XX DT 21-FEB-2000 (first entry)  
 XX  
 XX DE Human heparanase protein sequence.  
 XX  
 XX KW Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;  
 KW antidiabetic; immunomodulatory; anti-inflammatory; nephrotropic;  
 KW metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;  
 KW mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;  
 KW inflammation; haemorrhagic nephritis; nephrotic syndrome;  
 KW autoimmune disease; anticancer; kidney disease.  
 XX

```

OS Homo sapiens.
PN WO9957153-A1.
XX
PD 11-NOV-1999.
XX
PF 29-APR-1999; 99WO-US009255.
XX
PR 01-MAY-1998; 98US-00071739.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
PI Pecker I, Vlodavsky I, Friedman Y, Perets T;
XX
DR WPI; 2000-052944/04.
DR N-PSDB; AAZ33290.
XX
PT Heparanase-specific molecular probes useful for diagnosis and treatment,
PT e.g. of tumors, and for targeted drug delivery.
XX
PS Example; Page 81-82; 90pp; English.
XX
CC The present invention describes heparanase-specific molecular probes,
CC useful for methods of detecting heparanase in situ. The probes and anti-
CC heparanase antibodies are used to detect or quantify the expression of
CC heparanase, for diagnosis and monitoring of diseases (especially
CC metastasis), for treatment of heparanase-associated diseases (e.g.
CC tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,
CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its
CC metastases) derived from liver, prostate, bladder, breast, ovary, cervix,
CC colon, skin, intestine, stomach, uterus and pancreas, kidney disease,
CC diabetes and inflammation, haemorrhagic nephritis, nephrotic syndrome,
CC sepsis and inflammatory or autoimmune disease), for targeted drug
CC delivery (e.g. of anticancer agents) and as research reagents. The
CC present sequence represents human heparanase, which is used in the
CC exemplification of the present invention
XX
SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 3; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGETSSAY 15
Db |||||
334 RPKKKVWLGETSSAY 348

RESULT 32
AAY97635
ID AAY97635 standard; protein; 543 AA.
XX
AC AAY97635;
XX
DT 20-APR-2001 (first entry)
XX
DE Human heparanase protein sequence.
XX
KW Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW gene therapy; human.
XX
OS Homo sapiens.
XX
PN WO200100643-A2.
XX
PD 04-JAN-2001.
XX
PF 19-JUN-2000; 2000WO-IL000358.
XX

Query Match 100.0%; Score 82; DB 4; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKKVWLGETSSAY 15
Db |||||
334 RPKKKVWLGETSSAY 348

RESULT 33
AAB86206
ID AAB86206 standard; protein; 543 AA.
XX
AC AAB86206;
XX
DT 24-AUG-2001 (first entry)
XX
DE Human heparanase inhibitor protein.
XX
KW Heparanase; inhibitor; cardiac insufficiency; cardiant; nephrotropic;
KW hepatotropic; veterinary medicine; congestive heart failure; dyspnoea;
KW primary cardiomyopathy; peripheral odema; pulmonary congestion;
KW hepatic congestion; hydrothorax; ascite; nocturia; human.
XX
OS Homo sapiens.
XX
PN DE19955803-A1.
XX
PD 23-MAY-2001.
XX
PF 19-NOV-1999; 99DE-01055803.
XX
PR 19-NOV-1999; 99DE-01055803.
XX
PA (KNOL ) KNOLL AG.
XX
PI Herr D, Hahn A, Laux V;
XX
DR WPI; 2001-368371/39.
DR N-PSDB; AAH20940.
XX
PT Treatment or prevention of cardiac insufficiency and related conditions,
PT e.g. pulmonary congestion and dyspnoea, comprises administration of
PT heparanase inhibitor.
XX
PS Disclosure; Page 11-13; 16pp; German.
XX
CC This invention describes a novel heparanase inhibitor which can be used
CC for the treatment or prevention of cardiac insufficiency and associated
CC indications, symptoms and/or malfunctions. The heparanase inhibitor of

```



CC the invention has cardiant, nephrotropic and hepatotropic activity. The  
 CC products of the invention can be used in human and veterinary medicine,  
 CC for the treatment or prevention of congestive heart failure e.g. primary  
 CC cardiomyopathy. Associated conditions treated or prevented with the  
 CC inhibitor are especially peripheral odemas, pulmonary and hepatic  
 CC congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g.  
 CC nocturia can also be treated. This sequence represents the human  
 CC heparanase protein described in the method of the invention  
 XX  
 SQ

XX Sequence 543 AA;

Query Match 100.0%; Score 82; DB 4; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 0.00011;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPCKKVLGETSSAY 15  
 |||||  
 Db 334 RPCKKVLGETSSAY 348

RESULT 34

ABB07813  
 ID ABB07813 standard; protein; 543 AA.

XX ABB07813;

DT 03-JUL-2002 (first entry)

XX Human heparanase sequence.

XX Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;  
 KW anti-protozoan; neuroprotective; heparin; human.  
 XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1. .35

FT /note= "signal peptide"

FT Protein 36. .543

FT /note= "mature protein"

XX US2002034810-A1.

XX 21-MAR-2002.

XX 16-AUG-2001; 2001US-00930218.

XX 20-SEP-2000; 2000US-00666390.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia B;

XX WPI; 2002-338926/37.

XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful  
 PT to treat various heparin-related disorders and the signal peptide is  
 PT useful in production of membrane-targeted or secreted recombinant  
 PT proteins.

XX Disclosure; Fig 1a; 39pp; English.

XX The invention relates to an isolated avian and reptile nucleic acid,  
 CC encoding a polypeptide with heparanase catalytic activity. The signal  
 CC peptide of the nucleic acid can be used to express membrane-associated or  
 CC secreted proteins in heterologous expression systems. The encoded  
 CC polypeptides can be used to prevent tumour angiogenesis, metastasis and  
 CC invasion, and to intervene with pathologies associated with impaired  
 CC heparin-binding growth factors, cellular responses to heparin-binding  
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,  
 CC cellular susceptibility to viral, protozoa and bacterial infections or  
 CC disintegration of neurodegenerative plaques. The present sequence  
 CC represents a human heparanase protein sequence used in similarity studies

XX Sequence 543 AA;  
 SQ

Query Match 100.0%; Score 82; DB 5; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 0.00011;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPCKKVLGETSSAY 15  
 |||||  
 Db 334 RPCKKVLGETSSAY 348

RESULT 35

ADD18950

ID ADD18950 standard; protein; 543 AA.

XX ADD18950;

DT 15-JAN-2004 (first entry)

XX Human disease related protein SeqID439.

XX human; disease state; cytostatic; antiinflammatory; ophthalmological;  
 KW antiarteriosclerotic; vulnerary; gene therapy;  
 KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;  
 KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;  
 KW glucose transportation; catecholamine synthesis; iron transport;  
 KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;  
 KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;  
 KW inflammatory condition; wound healing.

XX Homo sapiens.

XX WO2003018621-A2.

XX 06-MAR-2003.

XX 23-AUG-2002; 2002WO-GB003892.

XX 23-AUG-2001; 2001GB-00020558.

XX 05-OCT-2001; 2001GB-00024037.

XX (OXFO-) OXFORD BIOMEDICA UK LTD.

XX Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;

XX WPI; 2003-290046/28.

XX N-PSDB; ADD18951.

XX New substantially purified polypeptide, useful for diagnosing or treating  
 PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion  
 PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or  
 PT wound healing.

XX Claim 25; SEQ ID NO 439; 424pp; English.

XX This invention relates to novel human genes and gene product which are  
 CC implicated in certain disease states. Compounds which modulate the  
 CC proteins of the invention may have cytostatic, antiinflammatory,  
 CC ophthalmological, antiarteriosclerotic or vulnerary activities. The  
 CC sequences of the invention may be useful for gene therapy. The invention  
 CC may be useful for diagnosing or treating a hypoxia-regulated condition,  
 CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,  
 CC erythropoiesis, or the biological response to hypoxia conditions  
 CC including processes such as glycolysis, gluconeogenesis, glucose  
 CC transportation, catecholamine synthesis, iron transport or nitric oxide  
 CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion  
 CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,  
 CC inflammatory conditions or wound healing. The present sequence is that of  
 CC a disease related protein of the invention.

XX Sequence 543 AA;

Query Match 100.0%; Score 82; DB 7; Length 543;  
Best Local Similarity 100.0%; Pred. No. 0.00011; Mismatches 0; Indels 0; Gaps 0;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPQKKVWLGETSSAY 15  
Db 334 RPQKKVWLGETSSAY 348  
|||||

RESULT 36  
ADG88800  
ID ADG88800 standard; protein; 543 AA.

AC ADG88800;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Human hpa protein.

KW Wound healing; heparanase; ulcer; burn; laceration; surgical incision;  
KW necrosis; pressure wound; diabetic ulcer; angiogenesis; human; therapy.  
XX  
OS Homo sapiens.

PN US2003161823-A1.  
XX  
PD 28-AUG-2003.

XX 14-JAN-2003; 2003US-00341582.

XX 31-AUG-1998; 98WO-US017954.  
PR 01-MAR-1999; 99US-00258892.  
PR 06-FEB-2001; 2001US-00776874.  
PR 05-SEP-2001; 2001WO-IL000830.  
PR 19-NOV-2001; 2001US-00988113.

XX (ILAN/) ILAN N.

PA (VLOD/) VLODAVSKY I.

PA (YACO/) YACOBY-ZEEVI O.

PA (PECK/) PECKER I.

PA (FEIN/) FEINSTEIN E.

XX Ilan N, Vlodavsky I, Yacoby-Zeevi O, Pecker I, Feinstein E;  
PI WPI; 2003-897910/82.

XX N-PSDB; ADG88799, ADG88801, ADG88832.  
XX

PT Composition for treating a wound comprising recombinant heparanase is  
PT useful to induce or accelerate wound healing and induce or accelerate  
PT angiogenesis.

PS Claim 2; SEQ ID NO 10; 143pp; English.

XX The present invention relates to methods and compositions for inducing  
CC and/or accelerating wound healing via the catalytic activity of  
CC heparanase. The invention is used to induce or accelerate a healing  
CC process, particularly of an ulcer, burn, laceration, surgical incision,  
CC necrosis, pressure wound, diabetic ulcer and to induce or accelerate  
CC angiogenesis. The present sequence is human hpa protein.

XX Sequence 543 AA;

Query Match 100.0%; Score 82; DB 7; Length 543;  
Best Local Similarity 100.0%; Pred. No. 0.00011; Mismatches 0; Indels 0; Gaps 0;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPQKKVWLGETSSAY 15  
Db 334 RPQKKVWLGETSSAY 348  
|||||

RESULT 37  
ADL16379

ID ADL16379 standard; protein; 543 AA.  
XX AC ADL16379;

XX 06-MAY-2004 (first entry)

XX Human heparanase partial protein.

XX Human; heparanase; heparanase-dependent cancer; cancer;  
KW autoimmune reaction; inflammation; chromosome 4; enzyme.

XX Homo sapiens.

XX US2003236215-A1.

XX 25-DEC-2003.

XX 09-JUN-2003; 2003US-00456573.

XX 31-AUG-1998; 98WO-US017954.

PR 01-MAR-1999; 99US-00258892.

PR 08-NOV-1999; 99US-00435739.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX Pecker I, Vlodavsky I, Feinstein E;

XX WPI; 2004-070610/07.

XX New antisense oligonucleotide hybridizable with a polynucleotide encoding  
PT a polypeptide with heparanase activity, useful for treating diseases such  
PT as cancer and autoimmune disorders.

XX Claim 3; SEQ ID NO 10; 108pp; English.

XX The invention relates to an antisense oligonucleotide (ASO) comprising a  
CC polynucleotide or a polynucleotide analogue of at least 10 bases being  
CC hybridisable in vivo , under physiological conditions, with a portion of  
CC a polynucleotide strand encoding a polypeptide having heparanase  
CC catalytic activity. Also included are a method of in vivo downregulating  
CC heparanase activity (comprising administering the ASO in vivo), a method  
CC of treating a subject suffering from a pathological condition  
CC (characterised by heparanase activity, comprising administering ASO to  
CC the subject), a pharmaceutical composition comprising the ASO and a  
CC carrier, an antisense nucleic acid construct (comprising a promoter  
CC sequence and a polynucleotide sequence directing the synthesis of an  
CC antisense RNA sequence of at least 10 bases being hybridisable in vivo ,  
CC under physiological conditions, with a polynucleotide strand encoding a  
CC polypeptide having heparanase catalytic activity), a method of in vivo  
CC downregulating heparanase activity (comprising administering in vivo the  
CC antisense nucleic acid construct), a pharmaceutical composition  
CC comprising the antisense nucleic acid construct and a carrier, and an  
CC antisense oligonucleotide comprising a polynucleotide or a polynucleotide  
CC analogue of at least 10 bases being hybridisable in vivo , under  
CC physiological conditions, with a portion of a polynucleotide strand being  
CC characterised by forming at least a portion of an untranslated region  
CC (UTR) for a polynucleotide strand encoding a polypeptide having  
CC heparanase catalytic activity. The methods and compositions of the  
CC present invention are useful for the prevention and/or treatment of  
CC diseases or conditions associated with aberrant heparanase activity, such  
CC as heparanase-dependent cancer, cancer, autoimmune reaction and  
CC inflammation. The gene for human heparanase is located on chromosome 4.  
CC The present sequence is a human heparanase protein.

XX Sequence 543 AA;

Query Match 100.0%; Score 82; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPQKKVWLGETSSAY 15  
|||||

```

Db      334  RPKKVLGETSSAY 348

RESULT 38
ADK52086
ID  ADK52086 standard; protein; 543 AA.
XX
AC  ADK52086;
XX
XX  20-MAY-2004 (first entry)
XX
XX  Human atopic dermatitis/psoriasis-associated protein #1.
XX
KW  Human; atopic dermatitis; psoriasis; dermatological; anti-inflammatory;
KW  antipsoriatic; rash.
XX
OS  Homo sapiens.
XX
XX  W02004016785-A1.
XX
XX  26-FEB-2004.
XX
XX  06-AUG-2003; 2003WO-JP009999.
XX
XX  06-AUG-2002; 2002JP-00229319.
XX  14-MAY-2003; 2003JP-00136544.
XX
XX  (GENO-) GENOX RES INC.
XX  (UJGU-) UNIV JUNTENDO.
XX
XX  Itoh M, Ogawa K, Shinagawa A, Sudo H, Ogawa H, Ra C;
XX  Mitsuishi K;
XX
XX  WPI; 2004-214514/20.
XX  N-PSDB; ADK51968.
XX
XX  Detecting atopic dermatitis or psoriasis comprises assaying levels of
XX  expression of an indicator gene at a rash site and non-rash site of a
XX  person with atopic dermatitis or psoriasis.
XX
XX  Example 2; SEQ ID NO 119; 484pp; Japanese.
XX
XX  The invention relates to detecting atopic dermatitis or psoriasis
XX  comprising assaying the levels of expression of an indicator gene at a
XX  rash site and non-rash site of a person with atopic dermatitis or
XX  psoriasis, comparing these levels with those of a healthy person, and
XX  determining that if the levels of indicators are higher or lower, then
XX  this indicates the disease. Also included are a reagent for detecting
XX  atopic dermatitis or psoriasis, a kit for screening for treatments, a
XX  transgenic non human vertebrate animal models for the diseases, an agent
XX  for inducing the diseases in mice and a DNA chip for assaying for the
XX  indicator genes. The method is used for treatment, detection and animal
XX  models for research of atopic dermatitis and psoriasis. The present
XX  sequence is a protein encoded by an indicator gene of the invention.
XX
XX  Sequence 543 AA;
XX
XX  Query Match      100.0%; Score 82; DB 8; Length 543;
XX  Best Local Similarity 100.0%; Pred. No. 0.00011;
XX  Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  QY      1  RPKKVLGETSSAY 15
XX          |||||
XX  Db      334  RPKKVLGETSSAY 348
XX
XX  RESULT 39
XX  ADM48716
XX  ID  ADM48716 standard; protein; 543 AA.
XX
XX  AC  ADM48716;
XX
XX  03-JUN-2004 (first entry)
XX
XX  Human hpa protein #2.
XX
XX  DE
XX
XX  Human hpa protein #1.
XX
XX  Transgenic animal; heparanase; cancer; viral infection; restenosis;
XX  neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;
XX  human.
XX
XX  Homo sapiens.
XX
XX  US2003217375-A1.
XX
XX  20-NOV-2003.
XX
XX  24-FEB-2003; 2003US-00371218.
XX
XX  31-AUG-1998; 98WO-US017954.
XX  01-MAR-1999; 99US-00258892.
XX  06-FEB-2001; 2001US-00776874.
XX  19-NOV-2001; 2001US-00988113.
XX
XX  (ZCHA/) ZCHARIA E.
XX  (VL0D/) VLODAVSKY I.
XX  (METZ/) METZGER S.
XX  (PECK/) PECKER I.
XX  (ILAN/) ILAN N.
XX  (CHAJ/) CHAJEK-SHAUL T.
XX  (GOLD/) GOLDSCHMIDT O.
XX
XX  Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;
XX  Chajek-Shaul T, Goldshmidt O;
XX
XX  WPI; 2004-021918/02.
XX  N-PSDB; ADM48715, ADM48717.
XX
XX  New transgenic non-human animal expressing heparinase, useful as models
XX  for human disease, such as cancers, viral infection, neurodegenerative
XX  diseases, restenosis, atherosclerosis and pulmonary disorders.
XX
XX  Example 1; SEQ ID NO 10; 106pp; English.
XX
XX  The present invention relates to a transgenic non-human animal whose
XX  genome comprises an exogenous polynucleotide sequence, including a
XX  promoter active in tissues of the non-human, a region encoding a human
XX  heparanase, where the promoter and the region encoding human heparanase
XX  are operably linked in the exogenous polynucleotide such that human
XX  heparanase is expressed in at least a portion of the cells of the non-
XX  human animal. The methods and compositions of the present invention are
XX  useful for the production of transgenic animals expressing heparanase, to
XX  be used as models for human diseases such as cancers, viral infection,
XX  restenosis, neurodegenerative diseases, atherosclerosis and pulmonary
XX  disorders. The present sequence is human hpa protein used in the
XX  exemplification of the invention.
XX
XX  Sequence 543 AA;
XX
XX  Query Match      100.0%; Score 82; DB 8; Length 543;
XX  Best Local Similarity 100.0%; Pred. No. 0.00011;
XX  Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  QY      1  RPKKVLGETSSAY 15
XX          |||||
XX  Db      334  RPKKVLGETSSAY 348
XX
XX  RESULT 40
XX  ADM48759
XX  ID  ADM48759 standard; protein; 543 AA.
XX
XX  AC  ADM48759;
XX
XX  03-JUN-2004 (first entry)
XX
XX  Human hpa protein #2.
XX
XX  DE

```

XX Transgenic animal; heparanase; cancer; viral infection; restenosis;  
KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;  
KW human.  
XX Homo sapiens.  
XX OS  
XX US2003217375-A1.  
XX PD  
XX 20-NOV-2003.  
XX  
XX 24-FEB-2003; 2003US-00371218.  
XX  
XX 31-AUG-1998; 98WO-US017954.  
XX 01-MAR-1999; 99US-00258892.  
XX 06-FEB-2001; 2001US-00776874.  
XX 19-NOV-2001; 2001US-00988113.  
XX  
XX (ZCHA/) ZCHARIA E.  
XX (VLOD/) VLODAVSKY I.  
XX (METZ/) METZGER S.  
XX (PECK/) PECKER I.  
XX (ILAN/) ILAN N.  
XX (CHAJ/) CHAJEK-SHAUL T.  
XX (GOLD/) GOLDSHMIDT O.  
XX  
XX Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;  
PI Chajek-Shaul T, Goldshmidt O;  
XX  
XX WPI: 2004-021918/02.  
XX N-PSDB; ADM48748.  
XX  
XX New transgenic non-human animal expressing heparinase, useful as models  
PT for human disease, such as cancers, viral infection, neurodegenerative  
PT diseases, restenosis, atherosclerosis and pulmonary disorders.  
XX  
XX Example 10; Fig 16; 106pp; English.  
XX  
XX The present invention relates to a transgenic non-human animal whose  
CC genome comprises an exogenous polynucleotide sequence, including a  
CC promoter active in tissues of the non-human, a region encoding a human  
CC heparanase, where the promoter and the region encoding human heparanase  
CC are operably linked in the exogenous polynucleotide such that human  
CC heparanase is expressed in at least a portion of the cells of the non-  
CC human animal. The methods and compositions of the present invention are  
CC useful for the production of transgenic animals expressing heparanase, to  
CC be used as models for human diseases such as cancers, viral infection,  
CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary  
CC disorders. The present sequence is human hpa protein used in the  
CC exemplification of the invention.  
XX  
XX Sequence 543 AA;  
XX  
XX Query Match 100.0%; Score 82; DB 8; Length 543;  
XX Best Local Similarity 100.0%; Pred. No. 0.00011;  
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1 RPKKKVWLGETSSAY 15  
XX |||||||||||  
XX 334 RPKKKVWLGETSSAY 348  
XX  
XX RESULT 41  
XX ADN05074  
XX ID ADN05074 standard; protein; 543 AA.  
XX  
XX ADN05074;  
XX  
XX 01-JUL-2004 (first entry)  
XX  
XX Antipsoriatic protein sequence #716.  
XX  
XX antipsoriatic; gene therapy; psoriasis; diagnosis.  
KW

XX Homo sapiens.  
XX WO2004028479-A2.  
XX  
XX 08-APR-2004.  
XX  
XX 25-SEP-2003; 2003WO-US030907.  
XX  
XX 25-SEP-2002; 2002US-0414006P.  
XX  
XX (GETH ) GENENTECH INC.  
XX  
XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;  
XX Wu TD;  
XX  
XX WPI; 2004-305105/28.  
XX N-PSDB; ADN05073.  
XX  
XX New PRO nucleic acid or polypeptide, useful for preparing a  
PT pharmaceutical composition for diagnosing or treating psoriasis in a  
PT mammal.  
XX  
XX Claim 9; SEQ ID NO 1468; 3069pp; English.  
XX  
XX The invention relates to novel polynucleotide and polypeptides for  
CC treating psoriasis or a sequence having at least 80% identity to the  
CC above sequences. The nucleic acid is useful for preparing a composition  
CC for diagnosing or treating psoriasis in a mammal. This sequence  
CC corresponds to one of the polypeptides of the invention.  
XX  
XX Sequence 543 AA;  
XX  
XX Query Match 100.0%; Score 82; DB 8; Length 543;  
XX Best Local Similarity 100.0%; Pred. No. 0.00011;  
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1 RPKKKVWLGETSSAY 15  
XX |||||||||||  
XX 334 RPKKKVWLGETSSAY 348  
XX  
XX RESULT 42  
XX ADN04902  
XX ID ADN04902 standard; protein; 543 AA.  
XX  
XX ADN04902;  
XX  
XX 01-JUL-2004 (first entry)  
XX  
XX Antipsoriatic protein sequence #631.  
XX  
XX antipsoriatic; gene therapy; psoriasis; diagnosis.  
KW  
XX Homo sapiens.  
XX WO2004028479-A2.  
XX  
XX 08-APR-2004.  
XX  
XX 25-SEP-2003; 2003WO-US030907.  
XX  
XX 25-SEP-2002; 2002US-0414006P.  
XX  
XX (GETH ) GENENTECH INC.  
XX  
XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;  
XX Wu TD;  
XX  
XX WPI; 2004-305105/28.  
XX N-PSDB; ADN04901.  
XX  
XX New PRO nucleic acid or polypeptide, useful for preparing a

PT pharmaceutical composition for diagnosing or treating psoriasis in a  
PT mammal.  
XX  
PS Claim 9; SEQ ID NO 1296; 3069pp; English.  
XX  
XX The invention relates to novel polynucleotide and polypeptides for  
CC treating psoriasis or a sequence having at least 80% identity to the  
CC above sequences. The nucleic acid is useful for preparing a composition  
CC for diagnosing or treating psoriasis in a mammal. This sequence  
CC corresponds to one of the polypeptides of the invention.  
XX  
XX Sequence 543 AA;  
SQ  
Query Match 100.0%; Score 82; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RPQKKVWLGETSSAY 15  
DB 334 RPQKKVWLGETSSAY 348  
|||||  
RESULT 43  
AD063831  
ID AD063831 standard; protein; 543 AA.  
XX  
AC AD063831;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE Human heparanase mutant E378A.  
XX  
KW Human; heparanase; heparanase-derived protein; heparanase mutant;  
KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;  
KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;  
KW vascular graft; platelet adhesion; platelet aggregation;  
KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
KW vulnery; mutant; mutein; enzyme.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Active-site 225 /note= "Active site proton donor"  
FT Active-site 343 /note= "Active site nucleophile"  
FT Misc-difference 378 /note= "Ala replaces wild-type Glu"  
FT  
FT  
FT  
FN WO2004048558-A2.  
XX  
XX 10-JUN-2004.  
XX  
XX 24-NOV-2003; 2003WO-IL000989.  
XX  
XX 24-NOV-2002; 2002IL-00153059.  
XX  
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
XX  
XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
XX WPI; 2004-450373/42.  
XX  
XX New nucleic acid construct comprising heparanase-derived polypeptide,  
PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
PT thrombasthenia, or Bernard-Soulier syndrome.  
XX  
XX Example 4; Page; 128pp; English.  
XX  
XX The invention relates to nucleic acid constructs comprising a nucleic  
CC acid encoding a heparanase-derived protein which lacks heparanase

CC endoglycosidase catalytic activity but which retains its cell-cell and  
CC cell-matrix adhesion properties. The constructs of the invention  
CC optionally further comprise operably linked regulatory elements. The  
CC invention also relates to the heparanase-derived proteins and host cells  
CC comprising the nucleic acid constructs of the invention. The heparanase-  
CC derived proteins are especially mutants of human heparanase in which the  
CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
CC are replaced with Ala (AD063822-AD063824), and the proteins may  
CC optionally further comprise an avian heparanase signal peptide (AD063825-  
CC AD063827). The heparanase-derived protein, nucleic acid construct and  
CC host cells are useful in preparing a tissue sealant composition for  
CC sealing injuries, reducing the loss of blood, accelerating the healing  
CC and homeostasis of an injury, accelerating blood vessel endothelium  
CC formation or the endothelialisation of vascular grafts, accelerating the  
CC adhesive activity of mammalian cells, and accelerating the adhesion and  
CC aggregation of platelets. They may also be used in the treatment of  
CC disorders associated with adhesion deficiency such as LAD (leukocyte  
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
CC cells of the invention may additionally be to screen for modulators of  
CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
CC synthetic vascular graft comprising a tube made of a biocompatible  
CC material lined with the cells. The present sequence represents a human  
CC heparanase mutant E378A created in an example of the invention which  
CC retains its heparanase catalytic activity. The present sequence is not  
CC shown in the invention, but is derived from the protein sequence of  
CC GenBank accession number AF144325 and the information provided on page  
CC 70.  
XX  
XX Sequence 543 AA;  
SQ  
Query Match 100.0%; Score 82; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RPQKKVWLGETSSAY 15  
DB 334 RPQKKVWLGETSSAY 348  
|||||  
RESULT 44  
AD063832  
ID AD063832 standard; protein; 543 AA.  
XX  
AC AD063832;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE Human heparanase mutant E396A.  
XX  
KW Human; heparanase; heparanase-derived protein; heparanase mutant;  
KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;  
KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;  
KW vascular graft; platelet adhesion; platelet aggregation;  
KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
KW vulnery; mutant; mutein; enzyme.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Active-site 225 /note= "Active site proton donor"  
FT Active-site 343 /note= "Active site nucleophile"  
FT Misc-difference 396 /note= "Ala replaces wild-type Glu"  
FT  
FT  
FT  
FN WO2004048558-A2.  
XX  
XX 10-JUN-2004.  
XX



ADQ80372  
 ID ADQ80372 standard; protein; 543 AA.  
 XX  
 AC ADQ80372;  
 XX  
 DT 21-OCT-2004 (first entry)  
 XX  
 DE Heparanase protein.  
 XX  
 KW cytostatic; epidermal growth factor receptor modulator; identification;  
 KW therapeutic response; cancer; EGFR; biomarker.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004063709-A2.  
 XX  
 PD 29-JUL-2004.  
 XX  
 XX 08-JAN-2004; 2004WO-US000368.  
 PF  
 XX 08-JAN-2003; 2003US-0438735P.  
 PR  
 XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
 PA  
 XX Amler LC, Januario T;  
 PI  
 XX WPI; 2004-544114/52.  
 DR  
 DR N-PSDB; ADQ80253.  
 XX  
 XX Identifying a mammal that will respond therapeutically to a method of  
 PT treating cancer comprises comparing the level of a biomarker in a mammal  
 PT before and after exposure to an epidermal growth factor receptor (EGFR)  
 PT modulator.  
 XX  
 XX Disclosure; SEQ ID NO 144; 520pp; English.  
 PS  
 XX  
 CC The invention relates to a method of identifying a mammal that will  
 CC respond therapeutically to a method of treating cancer by administering  
 CC an epidermal growth factor receptor (EGFR) modulator by comparing the  
 CC level of a biomarker in a mammal before and after exposure to an EGFR  
 CC modulator. The method comprises: (a) measuring, in the mammal, the level  
 CC of at least one biomarker identified in the specification; (b) exposing  
 CC the mammal to the EGFR modulator; and (c) measuring in the mammal the  
 CC level of the biomarker, where a difference in the level in step (c)  
 CC compared to step (a) indicates that the mammal will respond  
 CC therapeutically to the method of treating cancer. The method and  
 CC biomarkers are useful for identifying a mammal that will respond  
 CC therapeutically to a method of treating cancer by administering an  
 CC epidermal growth factor receptor (EGFR) modulator. This sequence  
 CC corresponds to one of the biomarkers whose levels of expression is  
 CC measured in the method of the invention.  
 XX  
 SQ Sequence 543 AA;  
 Query Match 100.0%; Score 82; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 0.00011;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPKGKWLGETSSAY 15  
 Db 334 RPKGKWLGETSSAY 348  
 |||||  
 RESULT 47  
 ADR88210  
 ID ADR88210 standard; protein; 543 AA.  
 XX  
 AC ADR88210;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 XX Human preproheparanase.  
 DE  
 XX

KW Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;  
 KW autoimmune disorder; cancer; angiogenesis; metastatic disease;  
 KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;  
 KW hematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;  
 KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;  
 KW human; heparanase; enzyme.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..35  
 FT Protein /label= Signal\_peptide  
 FT 36..543  
 FT /label= Mature\_heparanase  
 FT Region 36..109  
 FT /note= "8 KDa subunit of mature heparanase dimer"  
 FT Domain 89..107  
 FT /note= Functional peptide epitope  
 FT Region 158..543  
 FT /note= "45 KDa subunit of mature heparanase dimer"  
 FT Domain 219..233  
 FT /note= Functional peptide epitope  
 FT Active-site 225  
 FT /note= "Active site residue"  
 FT Binding-site 258..266  
 FT /note= "Putative heparin binding domain"  
 FT Domain 294..307  
 FT /note= Functional peptide epitope  
 FT Domain 334..348  
 FT /note= Functional peptide epitope  
 FT Active-site 343  
 FT /note= "Active site residue"  
 FT Binding-site 414..420  
 FT /note= "Putative heparin binding domain"  
 FT Domain 437..446  
 FT /note= Functional peptide epitope  
 FT  
 XX US2004170631-A1.  
 PN  
 XX 02-SEP-2004.  
 PD  
 XX 28-NOV-2003; 2003US-00722502.  
 PF  
 XX 02-SEP-1997; 97US-00922170.  
 PR 01-MAY-1998; 98US-00071739.  
 PR 04-NOV-1998; 98US-00186200.  
 PR 19-FEB-2003; 2003US-00368044.  
 PR 22-AUG-2003; 2003US-00645659.  
 XX  
 XX (YACO/) YACOBY-ZEEVI O.  
 PA (PERE/) PERETZ T.  
 PA (MIRO/) MIRON D.  
 PA (SHLO/) SHLOMI Y.  
 PA (PECK/) PECKER I.  
 PA (AYAL/) AYAL-HERSHKOVITZ M.  
 PA (FEIN/) FEINSTEIN E.  
 PA (VGEL/) VAN GELDER J M.  
 PA (VLOD/) VLODAVSKY I.  
 PA (FRIE/) FRIEDMANN Y.  
 XX  
 XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
 PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;  
 PI Friedmann Y;  
 XX  
 DR WPI; 2004-625084/60.  
 XX  
 XX Targeted drug delivery to a heparanase-expressing tissue of a patient,  
 PT useful for treating heparanase-associated conditions such as inflammation  
 PT or cancer, comprises administering a drug and an anti-heparanase antibody  
 PT complex.  
 PT  
 PS Claim 2; SEQ ID NO 4; 58pp; English.  
 XX





FT Binding-site 426...433  
XX /note= "Putative heparin binding site"  
PN US2004213789-A1.  
XX  
PD 28-OCT-2004.  
XX  
XX 22-AUG-2003; 2003US-00645659.  
XX  
XX 02-SEP-1997; 97US-00922170.  
PR 01-MAY-1998; 98US-00071739.  
PR 04-NOV-1998; 98US-00186200.  
PR 19-FEB-2003; 2003US-00368044.  
XX  
XX (YACO/) YACOBY-ZEEVI O.  
PA (PERE/) PERETZ T.  
PA (MIRO/) MIRON D.  
PA (SHLO/) SHLOMI Y.  
PA (PECK/) PECKER I.  
PA (AYAL/) AYAL-HERSHKOVITZ M.  
PA (FEIN/) FEINSTEIN E.  
PA (GELD/) GELDER J M V.  
PA (VLOB/) VLODAVSKY I.  
PA (FRIE/) FRIEDMANN Y.  
XX  
PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;  
PI Friedmann Y;  
XX  
XX WPI; 2004-774790/76.  
XX  
XX New neutralizing monoclonal anti-heparanase antibodies, useful for  
PT detecting, treating or preventing cancer, inflammatory or autoimmune  
PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.  
XX  
XX Claim 5; SEQ ID NO 4; 68pp; English.  
XX  
XX The invention relates to an isolated antibody or antibody portion capable  
CC of specifically binding to or elicited by at least one epitope of a  
CC heparanase protein, where the heparanase protein is at least 60%  
CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and  
CC where at least one epitope comprises a sequence at least 70% homologous  
CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)  
CC a hybridoma cell line comprising a cell line for producing the monoclonal  
CC antibody, (b) a method for detecting, treating or preventing a  
CC pathological condition or a heparanase-related disorder or condition in a  
CC subject, (c) a method for monitoring the state of a heparanase-related  
CC disorder or condition in a subject, and (d) a pharmaceutical composition  
CC comprising the isolated anti-heparanase antibody or antibody portion and  
CC a pharmaceutical carrier. The antibody, methods, and composition are  
CC useful for detecting, treating, preventing or monitoring a pathological  
CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition  
CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,  
CC or prostate cancer), minor cell proliferation, invasion of circulating  
CC tumour cells, or a metastatic disease, or a heparanase-related disorder  
CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy  
CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or  
CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,  
CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell  
CC carcinoma) in a mammal. This sequence represents human heparanase.  
XX  
XX Sequence 543 AA;  
SQ  
Query Match 100.0%; Score 82; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RPCKKWLGETSSAY 15  
DB 334 RPCKKWLGETSSAY 348  
RESULT 50

ADY27036  
ID ADY27036 standard; protein; 543 AA.  
XX  
XX AC ADY27036;  
XX  
XX DT 05-MAY-2005 (first entry)  
XX  
XX DE Human heparanase protein.  
XX  
XX KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
KW neurological disease; viral infection; infection; cytostatic;  
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
KW protease; enzyme; enzyme purification.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO2005016227-A2.  
XX  
XX PD 24-FEB-2005.  
XX  
XX PF 12-AUG-2004; 2004WO-IL000744.  
XX  
XX PR 14-AUG-2003; 2003US-0494800P.  
PR 12-JAN-2004; 2004US-0535492P.  
XX  
XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.  
XX  
XX PI Van-Gelder JM, Miron D;  
XX  
XX DR WPI; 2005-182203/19.  
XX  
XX Regulating heparanase activity, useful for treating heparanase-associated  
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
PT neurological diseases or viral diseases) comprises modulating heparanase  
PT activation.  
XX  
XX Disclosure; SEQ ID NO 8; 21pp; English.  
XX  
XX The invention relates to a method of regulating heparanase activity in a  
CC tissue or regulating a biological process depending at least in part on  
CC heparanase activity comprising modulating heparanase activation. The  
CC invention also relates to methods of treating a heparanase- or heparin  
CC binding protein-associated disease or disorder in a subject, a  
CC pharmaceutical composition for use in the treatment of a heparanase-  
CC associated disease or disorder comprising a therapeutic amount of an  
CC agent capable of modulating heparanase activation and a pharmaceutical  
CC carrier or diluent, a method of identifying a protease activator of  
CC heparanase, a protease substrate mimetic comprising a peptide  
CC representing a subset or all substrate residues or cleavage sites of  
CC human heparanase or an equivalent non-human heparanase, a method of  
CC producing active heparanase and a method of modulating an adhesion  
CC activity of heparanase. The composition and methods are useful for  
CC modulating heparanase activation and for treating heparanase-associated  
CC diseases or disorders such as cancer, inflammation, cardiovascular  
CC diseases, neurological diseases or viral infections. This sequence  
CC represents a human heparanase protein used in the scope of the invention.  
XX  
XX Sequence 543 AA;  
SQ  
Query Match 100.0%; Score 82; DB 9; Length 543;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RPCKKWLGETSSAY 15  
DB 334 RPCKKWLGETSSAY 348  
Search completed: June 5, 2006, 12:41:01  
Job time : 136.384 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

QM protein - protein search, using sw model

Run on: June 5, 2006, 12:43:17 ; Search time 16.6438 Seconds

(without alignments)  
86.714 Million cell updates/sec

Title: US-10-645-659A-9

Perfect score: 82

Sequence: 1 RPKKVLGHTSSAY 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

1: PIR\_80.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	57.3	426	2 S44953	lmbF protein - Str
2	46	56.1	1180	1 NCECX5	exodeoxyribonuclea
3	46	56.1	1180	2 G85933	DNA helicase RecB
4	46	56.1	1180	2 E91088	DNA helicase RecB
5	45.5	55.5	1596	2 AG2501	hypothetical prote
6	45	54.9	177	2 D84828	AP2 domain transcr
7	45	54.9	562	2 AD0783	PTS system, fructo
8	44	53.7	336	2 E84594	AP2 domain transcr
9	44	53.7	563	2 E85855	PTS system, fructo
10	44	53.7	563	2 C91011	fructose-specific
11	44	53.7	563	2 A34962	phosphotransferase
12	43	52.4	249	2 A83664	tRNA/rRNA methyltr
13	43	52.4	1022	2 T51257	calmodulin-binding
14	43	52.4	1022	2 T50928	calmodulin-binding
15	43	52.4	1181	2 AB0865	exonuclease V chai
16	42	51.2	283	2 C84828	AP2 domain transcr
17	42	51.2	366	2 C83034	probable oxidoredu
18	42	51.2	775	2 E70320	polyribonucleotide
19	42	51.2	1366	1 CGH2S	collagen alpha 2(I
20	41	50.0	225	2 T02433	DNA binding protei
21	41	50.0	306	2 AH2866	S-methyltransferas
22	41	50.0	306	2 E97643	mah protein limpor
23	41	50.0	420	2 S16444	hygromycin phospho
24	41	50.0	480	2 JC7506	heparanase protein
25	41	50.0	615	2 AH2248	proteinase limpor
26	41	50.0	1220	2 AD0125	exodeoxyribonuclea
27	40	48.8	122	2 AE3301	2-deoxy-D-gluconat
28	40	48.8	131	2 S64937	probable membrane
29	40	48.8	222	2 T52019	ethylene responsiv

30	40	48.8	239	2 G97553	sugar fermentation
31	40	48.8	239	2 A12773	sugar fermentation
32	40	48.8	243	2 B82979	hypothetical prote
33	40	48.8	276	2 E72623	probable autoantig
34	40	48.8	291	2 T02434	DNA binding protei
35	40	48.8	326	2 T09544	phosphoprotein pho
36	40	48.8	386	2 JC5191	contractile tail s
37	40	48.8	958	2 S73012	polyketide synthas
38	40	48.8	1014	2 F95501	hypothetical prote
39	40	48.8	1126	1 W9F012	125K protein - alf
40	40	48.8	1871	2 AB7204	polyketide synthas
41	39	47.6	115	2 AB2718	conserved hypothet
42	39	47.6	182	2 F97499	hypothetical prote
43	39	47.6	184	2 A86461	hypothetical prote
44	39	47.6	196	2 H64968	acetyl CoA acetyl
45	39	47.6	221	2 T17997	hypothetical prote
46	39	47.6	308	2 F64901	ABC-type transport
47	39	47.6	308	2 B85728	hypothetical prote
48	39	47.6	308	2 G90889	hypothetical prote
49	39	47.6	366	1 XNB7UG	UDPglucose-hexose-
50	39	47.6	379	2 B70579	probable cell divi
51	39	47.6	382	1 SUBSN	subtilisin (EC 3.4
52	39	47.6	410	2 F87004	arabinogalactan en
53	39	47.6	425	2 F70038	probable mannose-6
54	39	47.6	751	2 G01234	N-ethylmaleimide-s
55	39	47.6	752	2 S04235	vesicular fusion p
56	39	47.6	828	2 T06133	hypothetical prote
57	39	47.6	1197	2 D82696	hypothetical prote
58	39	47.6	1373	1 A43291	collagen alpha 2(I
59	39	47.6	1760	2 E86201	protein Fl2K11.4 [
60	39	47.6	4151	2 G70944	probable polyketid
61	38	46.3	136	2 T36624	hypothetical prote
62	38	46.3	143	2 S07234	collagen alpha 1(I
63	38	46.3	183	2 T14653	hypothetical prote
64	38	46.3	195	2 T14965	hypothetical prote
65	38	46.3	204	2 S49031	cadmium-induced pr
66	38	46.3	207	2 D82423	hypothetical prote
67	38	46.3	248	2 T07728	transcription fact
68	38	46.3	251	2 T01986	Tsai protein - com
69	38	46.3	281	2 T52189	ethylene responsiv
70	38	46.3	297	2 G70185	nevalonate kinase
71	38	46.3	323	2 F90479	(Phospho) mevalona
72	38	46.3	335	2 T51833	transcription fact
73	38	46.3	358	2 D96579	hypothetical prote
74	38	46.3	362	2 F75379	S-adenosylmethioni
75	38	46.3	378	2 S54056	probable membrane
76	38	46.3	437	2 S60957	transcription modu
77	38	46.3	444	2 C83336	probable cytochrom
78	38	46.3	454	1 NMIV	exo-alpha-sialidas
79	38	46.3	462	2 S10235	alpha-L-fucosidase
80	38	46.3	470	1 NMIVAK	exo-alpha-sialidas
81	38	46.3	470	1 NMIVXL	exo-alpha-sialidas
82	38	46.3	470	2 S04801	exo-alpha-sialidas
83	38	46.3	473	2 I50629	collagen - chicken
84	38	46.3	493	2 G83564	probable ATPase PA
85	38	46.3	520	2 S74497	hypothetical prote
86	38	46.3	603	2 J71444	probable EREBP-4 -
87	38	46.3	613	2 J77827	X-Pro aminopeptida
88	38	46.3	662	2 T04461	potassium channel
89	38	46.3	810	2 T10756	Nel-homolog protei
90	38	46.3	958	2 S41013	hypothetical prote
91	38	46.3	1010	2 T09499	ATP-dependent clp
92	38	46.3	1032	2 T23164	hypothetical prote
93	38	46.3	1035	2 T23165	hypothetical prote
94	38	46.3	1040	2 D88568	protein ZK57.3 [1
95	38	46.3	3140	1 GNV5RA	genome polyprotein
96	38	46.3	3140	2 S47508	genome polyprotein
97	37.5	45.7	279	2 AC3114	hypothetical prote
98	37.5	45.7	304	2 B98173	hypothetical prote
99	37.5	45.7	1004	2 T38074	hypothetical prote
100	37.5	45.7	1401	2 T17452	Werner syndrome pr
101	37.5	45.7	2581	2 AF2545	hypothetical prote
102	37	45.1	94	2 T15982	hypothetical prote

103	37	45.1	161	2	T07689	transcription fact	176	36	43.9	300	2	T52020	ethylene responsiv
104	37	45.1	196	2	H44701	ribosomal protein	177	36	43.9	306	2	B28170	reaction center pr
105	37	45.1	206	2	S75258	probable 3-demethy	178	36	43.9	310	2	F81878	probable prolyl am
106	37	45.1	258	2	A13450	ferredoxin-NADP re	179	36	43.9	317	2	B81141	proline iminopecti
107	37	45.1	270	2	C86670	hypothetical prote	180	36	43.9	310	2	T03304	probable phosphopr
108	37	45.1	303	2	T04541	hypothetical prote	181	36	43.9	317	2	D82742	dihydroxydipicolin
109	37	45.1	344	2	T44032	integral membrane	182	36	43.9	318	2	D82742	phosphoprotein pho
110	37	45.1	364	2	D75539	branched-chain ami	183	36	43.9	323	2	A32549	phosphoprotein pho
111	37	45.1	398	2	AF1903	hypothetical prote	184	36	43.9	325	2	T09547	phosphoprotein pho
112	37	45.1	408	2	G87646	molybdopterin bios	185	36	43.9	325	2	AH2555	hypothetical prote
113	37	45.1	422	2	D84403	dihydrocorotase lim	186	36	43.9	326	2	T09995	phosphoprotein pho
114	37	45.1	428	2	T24769	hypothetical prote	187	36	43.9	334	2	D84978	lipoprotein nlpD p
115	37	45.1	432	2	AC0161	serine-type D-Ala-	188	36	43.9	337	2	E86543	probable RAV-like
116	37	45.1	441	2	H82642	hypothetical prote	189	36	43.9	380	2	G89604	spore coat protein
117	37	45.1	443	1	BVB91	hypothetical prote	190	36	43.9	402	2	F81387	UDP-N-acetylmuramo
118	37	45.1	451	2	G01227	mevalonate kinase	191	36	43.9	407	1	KHRTD	cathepsin D (EC 3.
119	37	45.1	455	2	AD2049	recepin - human	192	36	43.9	407	2	T36649	probable transposa
120	37	45.1	468	1	NMIVAA	aminopeptidase P [	193	36	43.9	408	2	H95416	probable ROK famil
121	37	45.1	512	2	A55206	exo-alpha-sialidas	194	36	43.9	410	1	KHMSD	cathepsin D (EC 3.
122	37	45.1	521	2	T45608	sucrose/fructan hy	195	36	43.9	412	1	KHHUD	cathepsin D (EC 3.
123	37	45.1	521	2	T45608	hypothetical prote	196	36	43.9	416	2	D84829	hypothetical prote
124	37	45.1	579	2	A70954	hypothetical prote	197	36	43.9	442	2	JC5218	type I site-specif
125	37	45.1	592	2	G69252	aldehyde ferredoxi	198	36	43.9	447	2	S20711	neuraminidase - in
126	37	45.1	592	2	E89772	hypothetical prote	199	36	43.9	453	1	NMIV3	exo-alpha-sialidas
127	37	45.1	649	2	T33741	DNA-binding protei	200	36	43.9	466	2	S53713	nitric-oxide reduc
128	37	45.1	663	1	A69798	beta-galactosidase	201	36	43.9	470	1	NMIVEK	exo-alpha-sialidas
129	37	45.1	703	1	WMBET7	U17 protein - hum	202	36	43.9	498	2	H81782	adhesin MafB2 NMA2
130	37	45.1	722	2	B75608	GMC oxidoreductase	203	36	43.9	501	2	C82414	aminopeptidase VCA
131	37	45.1	735	2	JC5869	beta-glucosidase (	204	36	43.9	506	2	T12819	hypothetical prote
132	37	45.1	826	2	T19049	hypothetical prote	205	36	43.9	533	2	D86756	prophage p12 prote
133	37	45.1	825	2	H64083	biotin sulfoxide r	206	36	43.9	587	1	TVFVPR	protein-tyrosine k
134	37	45.1	841	2	AC2250	nitrogen assimilati	207	36	43.9	667	2	R86728	NADH dehydrogenase
135	37	45.1	903	2	C83044	Mg(2+) transport A	208	36	43.9	686	2	T12967	hypothetical prote
136	37	45.1	1556	1	A60988	saliva-interacting	209	36	43.9	711	2	C83922	ATP-dependent prot
137	37	45.1	1561	1	S06839	surface antigen sp	210	36	43.9	812	2	H86265	probable secreted
138	37	45.1	1585	2	S04729	hypothetical prote	211	36	43.9	824	2	T36818	alpha-L-arabinofur
139	37	45.1	1630	2	T40217	microtubule bindin	212	36	43.9	825	2	A59296	hypothetical prote
140	37	45.1	1690	2	T13030	calcium channel pr	213	36	43.9	872	2	AD2216	leucyl-tRNA synthet
141	36.5	44.5	2178	2	S29237	phosphoribosyl-AMP	214	36	43.9	915	2	S54485	CESI protein - yea
142	36.5	44.5	86	2	F34964	Ig heavy chain V-I	215	36	43.9	928	2	AF3516	Mg(2+) transport A
143	36.5	44.5	319	2	A70605	probable acyl-coad	216	36	43.9	950	2	F86286	hypothetical prote
144	36.5	44.5	331	2	T35809	hypothetical prote	217	36	43.9	1039	2	T30856	protein F2 - Strept
145	36.5	44.5	380	2	D90749	probable membrane	218	36	43.9	1249	2	AC1065	helicase related p
146	36.5	44.5	380	2	H85599	probable membrane	219	36	43.9	1301	2	S18118	alpha-amylase - Al
147	36.5	44.5	380	2	F64826	probable membrane	220	36	43.9	1404	2	F86470	probable retroelem
148	36.5	44.5	907	2	F84825	hypothetical prote	221	36	43.9	1442	2	C82898	DNA polymerase III
149	36	43.9	83	2	G82744	hypothetical prote	222	36	43.9	1643	2	T07961	myosin heavy chain
150	36	43.9	102	2	T31572	hypothetical prote	223	36	43.9	1707	2	A33526	collagen alpha 2(I
151	36	43.9	107	2	B97204	hypothetical prote	224	36	43.9	1759	2	T03725	replicase polyprot
152	36	43.9	118	1	B42959	14K hypothetical p	225	36	43.9	3519	2	S43048	polyketide synthas
153	36	43.9	118	2	H85889	hypothetical prote	226	36	43.9	3643	2	T36410	probable polyketid
154	36	43.9	118	2	E91045	hypothetical prote	227	35.5	43.3	110	2	T17913	hypothetical prote
155	36	43.9	118	2	T07905	low-carbon dioxide	228	35.5	43.3	250	2	S40405	MADS box protein o
156	36	43.9	133	2	D86175	hypothetical prote	229	35.5	43.3	327	2	AG3367	alcohol dehydrogen
157	36	43.9	143	2	S00718	ribosomal protein	230	35.5	43.3	384	2	G82976	probable rubredoxi
158	36	43.9	166	2	S28242	NADH2 dehydrogenas	231	35.5	43.3	472	2	AC3534	glu/asp-tRNA amido
159	36	43.9	197	1	RPBR16	repressor protein	232	35.5	43.3	546	1	T34480	serine-tRNA ligase
160	36	43.9	210	2	F72224	conserved hypothet	233	35.5	43.3	2347	1	TVHURS	kinase-related pro
161	36	43.9	223	2	F81611	ribosomal protein	234	35.5	43.3	2492	1	A44213	nonstructural poly
162	36	43.9	223	2	F86570	S3 ribosomal prote	235	35.5	43.3	3947	2	T52486	ferrichrome sidero
163	36	43.9	223	2	C72055	S3 ribosomal prote	236	35	42.7	88	2	AB3320	hypothetical prote
164	36	43.9	236	2	T02577	probable AP2 domai	237	35	42.7	94	2	A27691	methionine gamma-L
165	36	43.9	244	2	B86197	hypothetical prote	238	35	42.7	95	2	G37262	Ig heavy chain V r
166	36	43.9	247	2	T06721	hypothetical prote	239	35	42.7	104	2	E82360	cya protein VC012
167	36	43.9	248	1	G64839	ymcB protein - Esc	240	35	42.7	107	2	G83348	hypothetical prote
168	36	43.9	248	2	E90771	hypothetical prote	241	35	42.7	138	2	S15441	ribosomal protein
169	36	43.9	248	2	A85634	hypothetical prote	242	35	42.7	138	2	AC3337	hypothetical prote
170	36	43.9	257	2	G84712	hypothetical prote	243	35	42.7	141	2	A70556	probable mutator M
171	36	43.9	260	2	S74597	ABC-type transport	244	35	42.7	147	2	AC1367	conserved hypothet
172	36	43.9	273	1	R5R22	ribosomal protein	245	35	42.7	147	2	AD1736	conserved hypothet
173	36	43.9	273	1	R5ZM2	ribosomal protein	246	35	42.7	153	2	S47943	translation initia
174	36	43.9	284	2	B87649	hypothetical prote	247	35	42.7	160	2	A10760	probable propanedi
175	36	43.9	298	2	C86871	conserved hypothet	248	35	42.7	183	2	S20844	modulation protein

249 35 42.7 183 2 G95313  
250 35 42.7 186 1 AXPG  
251 35 42.7 188 2 S66001  
252 35 42.7 196 2 C84919  
253 35 42.7 198 2 E83211  
254 35 42.7 215 2 D81120  
255 35 42.7 216 2 D97119  
256 35 42.7 236 2 B84718  
257 35 42.7 245 2 E86168  
258 35 42.7 246 2 H95408  
259 35 42.7 248 2 S03888  
260 35 42.7 250 2 S41513  
261 35 42.7 253 2 AB3280  
262 35 42.7 255 1 VGBE2E  
263 35 42.7 257 2 B82837  
264 35 42.7 258 2 JH0472  
265 35 42.7 259 2 G85363  
266 35 42.7 260 2 S10016  
267 35 42.7 261 2 A50405  
268 35 42.7 263 2 JC5271  
269 35 42.7 266 2 A80665  
270 35 42.7 268 2 AG3322  
271 35 42.7 273 2 T47822  
272 35 42.7 277 2 T03927  
273 35 42.7 278 2 A75427  
274 35 42.7 281 2 AF0505  
275 35 42.7 290 2 JU0473  
276 35 42.7 291 2 D70729  
277 35 42.7 300 2 D70018  
278 35 42.7 306 2 A10441  
279 35 42.7 312 2 S53969  
280 35 42.7 314 2 AC2868  
281 35 42.7 314 2 G97644  
282 35 42.7 317 2 F72246  
283 35 42.7 317 2 AD3067  
284 35 42.7 319 2 AH2062  
285 35 42.7 321 2 S31088  
286 35 42.7 323 2 T09550  
287 35 42.7 326 2 T09548  
288 35 42.7 327 1 S13827  
289 35 42.7 327 2 A82241  
290 35 42.7 329 2 G71140  
291 35 42.7 329 2 T47795  
292 35 42.7 338 2 F98219  
293 35 42.7 338 2 B49642  
294 35 42.7 341 2 F69171  
295 35 42.7 349 2 E90453  
296 35 42.7 352 2 F96548  
297 35 42.7 370 2 A75371  
298 35 42.7 370 2 B72594  
299 35 42.7 382 2 AC3118  
300 35 42.7 391 2 AG2536  
301 35 42.7 398 2 JC4174  
302 35 42.7 399 2 B71140  
303 35 42.7 402 2 D98169  
304 35 42.7 421 2 T33811  
305 35 42.7 421 2 T42252  
306 35 42.7 425 2 T12473  
307 35 42.7 425 2 S14147  
308 35 42.7 435 2 T20819  
309 35 42.7 453 2 A48348  
310 35 42.7 454 2 A82952  
311 35 42.7 462 2 T17480  
312 35 42.7 466 2 AF0190  
313 35 42.7 469 1 NMIVH3  
314 35 42.7 469 1 NMIVN2  
315 35 42.7 486 1 COBYW2  
316 35 42.7 488 2 B47032  
317 35 42.7 488 2 A95301  
318 35 42.7 498 1 VGXPLA  
319 35 42.7 498 1 VGXPLC  
320 35 42.7 498 1 VGXPLM  
321 35 42.7 506 2 S75789

NodL Nod factor ac  
adrenodoxin precu  
conserved hypothet  
hypothetical prote  
hypothetical prote  
probable periplasm  
amidase from nicot  
hypothetical prote  
hypothetical prote  
probable short cha  
photosystem II oxy  
Brn-3c protein - m  
hypothetical membr  
glycoprotein E - e  
conserved hypothet  
apolipoprotein A-I  
hypothetical prote  
photosystem II oxy  
methionyl aminopep  
oxygen-evolving co  
conserved hypothet  
anhydro-N-acetylmu  
hypothetical prote  
DNA binding protei  
hypothetical prote  
probable exported  
ypua protein - bac  
probable peptide t  
sugar permease hom  
probable membrane  
hypothetical prote  
UDP-hexose transfe  
UDP-hexose transfe  
glyoxalase - Ther  
hypothetical prote  
aldo/keto reductas  
phosphoprotein pho  
phosphoprotein pho  
phosphoprotein pho  
phosphoprotein pho  
probable ABC trans  
hypothetical prote  
hypothetical prote  
hypothetical prote  
POU-domain protein  
conserved hypothet  
glycosyltransferas  
probable DNA-bindi  
threonine 3-dehydr  
hypothetical prote  
transcription regu  
hypothetical prote  
methionine gamma-1  
methionine prote  
hypothetical prote  
hypothetical prote  
polypeptide N-acet  
hypothetical prote  
multifunctional pu  
hypothetical prote  
genome polyprotein  
glucosamine-1-phos  
endo-xylanase homo  
probable transmemb  
exo-alpha-sialidas  
exo-alpha-sialidas  
whi2 protein - yea  
galactose-1-phosph  
hypothetical prote  
surface glycoprote  
surface glycoprote  
surface glycoprote  
6-aminohexanoate-c

322 35 42.7 513 2 T46143  
323 35 42.7 516 2 G70556  
324 35 42.7 518 2 T05277  
325 35 42.7 519 2 E87233  
326 35 42.7 527 2 T27572  
327 35 42.7 529 2 T45254  
328 35 42.7 537 2 T45634  
329 35 42.7 543 2 B82215  
330 35 42.7 548 2 T42617  
331 35 42.7 550 1 VGBEG5  
332 35 42.7 552 1 VGBEKD  
333 35 42.7 558 2 AG1236  
334 35 42.7 558 2 AB1599  
335 35 42.7 562 2 S25725  
336 35 42.7 566 2 AGU158  
337 35 42.7 578 2 B64012  
338 35 42.7 581 2 JC7086  
339 35 42.7 589 2 T12820  
340 35 42.7 595 2 H81044  
341 35 42.7 626 2 JQ2322  
342 35 42.7 630 2 A47398  
343 35 42.7 642 2 S55521  
344 35 42.7 651 2 T12083  
345 35 42.7 667 2 S56829  
346 35 42.7 678 2 S77215  
347 35 42.7 708 2 T19474  
348 35 42.7 720 1 HJBCD2  
349 35 42.7 720 2 AG0918  
350 35 42.7 720 2 G91221  
351 35 42.7 720 2 B86068  
352 35 42.7 722 2 F82482  
353 35 42.7 723 2 E82353  
354 35 42.7 747 2 T23607  
355 35 42.7 761 2 D70750  
356 35 42.7 781 2 AH3355  
357 35 42.7 784 2 AH2560  
358 35 42.7 795 2 T20609  
359 35 42.7 841 2 A90669  
360 35 42.7 841 2 D85519  
361 35 42.7 841 2 C64755  
362 35 42.7 910 2 A86782  
363 35 42.7 925 2 T29585  
364 35 42.7 962 2 D70661  
365 35 42.7 968 2 C82452  
366 35 42.7 994 2 AC3480  
367 35 42.7 1046 2 T42734  
368 35 42.7 1101 2 G70951  
369 35 42.7 1194 2 E88499  
370 35 42.7 1293 2 S24402  
371 35 42.7 1460 1 RNEY3L  
372 35 42.7 1530 2 AD1663  
373 35 42.7 1582 2 AC1153  
374 35 42.7 1728 2 T17466  
375 35 42.7 4976 2 T14165  
376 34.5 42.1 114 2 F0181  
377 34.5 42.1 116 2 S09962  
378 34.5 42.1 140 2 S09216  
379 34.5 42.1 288 2 S18990  
380 34.5 42.1 390 2 C70765  
381 34.5 42.1 396 2 D82584  
382 34.5 42.1 402 2 E70656  
383 34.5 42.1 491 2 B96739  
384 34.5 42.1 618 2 D88483  
385 34 41.5 49 2 I48681  
386 34 41.5 64 2 T07567  
387 34 41.5 72 2 A26084  
388 34 41.5 79 2 T18142  
389 34 41.5 86 2 T17313  
390 34 41.5 108 1 S77013  
391 34 41.5 118 2 AF0816  
392 34 41.5 129 1 MVECMT  
393 34 41.5 132 2 G85492  
394 34 41.5 132 2 G90641

steroid 22-alpha-h  
probable trpg prot  
dihydrofolate redu  
probable secreted  
hypothetical prote  
probable anthranil  
hypothetical prote  
methyl-accepting c  
probable envelope  
glycoprotein E pre  
glycoprotein E pre  
glycerol 3 phospho  
glycerol 3 phospho  
poly(3-hydroxyalka  
protein-Npi-phosph  
hypothetical prote  
FZD10 protein - hu  
hypothetical prote  
hemolysin activati  
starch synthase (E  
serotonin transpor  
beta-fructofuranos  
beta-fructofuranos  
probable membrane  
hypothetical prote  
hypothetical prote  
DNA helicase II (E  
DNA helicase II (i  
DNA-dependent ATPa  
DNA-dependent ATPa  
hypothetical prote  
DNA helicase II VC  
hypothetical prote  
probable ctcpa prot  
outer membrane pro  
hypothetical prote  
hypothetical prote  
probable enzyme (i  
yagk protein - Esc  
cation-transporin  
cytoplasmic linker  
probable ATP-depen  
protein K04G7.3 (i  
xeroderma pigmento  
DNA-directed RNA p  
glutamate synthase  
adhesin homolog lm  
rifamycin polyketi  
peptide synthetase  
ig heavy chain V r  
ig heavy chain V-D  
ig heavy chain pre  
probable acyl-CoA  
probable cobL - My  
penicillin binding  
hypothetical prote  
hypothetical prote  
protein let-721 (i  
185 kDa glycophosp  
hypothetical prote  
regulatory protein  
hypothetical prote  
hypothetical prote  
hypothetical prote  
conserved hypothet  
7,8-dihydro-8-oxog  
7,8-dihydro-8-oxog  
7,8-dihydro-8-oxog

395	34	41.5	162	2	H75435	hypothetical prote	468	34	41.5	381	2	AF3303	putrescine transpo
396	34	41.5	163	2	T36335	probable inorganic	469	34	41.5	382	2	F87464	conserved hypotet
397	34	41.5	165	2	D69784	hypothetical prote	470	34	41.5	384	2	B86660	amino acid amidohy
398	34	41.5	168	2	AH1176	B. subtilis regula	471	34	41.5	384	2	JC5206	hypothetical 43.0K
399	34	41.5	168	2	D84732	hypothetical prote	472	34	41.5	389	2	B82819	outer membrane pro
400	34	41.5	171	2	T00432	probable AP2 domai	473	34	41.5	393	2	A82664	conserved hypotet
401	34	41.5	194	2	H84771	probable AP2 domai	474	34	41.5	401	2	T35956	probable acyl-CoA
402	34	41.5	202	2	S30164	strong gravity str	475	34	41.5	401	2	A97446	hypothetical prote
403	34	41.5	203	2	T04311	lexA protein - Pse	476	34	41.5	408	2	A82162	hypothetical prote
404	34	41.5	204	2	S30165	repressor protein	477	34	41.5	410	2	S38238	hypothetical prote
405	34	41.5	207	2	T49897	transcription fact	478	34	41.5	414	2	A69907	cell wall-binding
406	34	41.5	208	2	T17082	NADH2 dehydrogenas	479	34	41.5	429	2	C84194	hypothetical prote
407	34	41.5	216	2	C87254	hypothetical prote	480	34	41.5	432	2	D95369	conserved hypotet
408	34	41.5	224	2	D83570	probable nucleotid	481	34	41.5	434	2	S37610	hypothetical metab
409	34	41.5	225	2	A85196	EREBP-2 protein [i	482	34	41.5	434	2	AC2833	hypothetical prote
410	34	41.5	225	2	T52011	ethylene responsiv	483	34	41.5	457	2	T39751	major facilitator
411	34	41.5	233	2	T02590	DNA binding protei	484	34	41.5	459	2	A82390	thiophen / furan o
412	34	41.5	233	2	T02590	transcription fact	485	34	41.5	459	2	S28025	light harvesting c
413	34	41.5	236	2	T02432	ethylene-responsiv	486	34	41.5	469	1	NM1V2	exo-alpha-sialidas
414	34	41.5	237	1	S11927	licheninase (EC 3.	487	34	41.5	472	2	B75501	glycogen synthase
415	34	41.5	238	2	AB3342	sugar fermentation	488	34	41.5	478	2	G75052	pyruvate kinase (E
416	34	41.5	238	2	S70195	kfiA protein - Bsc	489	34	41.5	478	2	F71171	probable pyruvate
417	34	41.5	240	2	H82289	conserved hypotet	490	34	41.5	481	2	H96529	hypothetical prote
418	34	41.5	243	2	T51989	ethylene responsiv	491	34	41.5	483	2	JC7179	acid phosphatase (
419	34	41.5	245	2	F70583	hypothetical prote	492	34	41.5	488	2	T30914	xylan 1,4-beta-xy
420	34	41.5	254	2	AI1117	hypothetical prote	493	34	41.5	491	2	T49304	hypothetical prote
421	34	41.5	254	2	AC1478	dehydrogenase/redu	494	34	41.5	494	2	A58721	sulfate adenylyltr
422	34	41.5	254	2	F90107	60S ribosomal prot	495	34	41.5	498	2	S27849	variant surface gl
423	34	41.5	255	2	G97227	protein containing	496	34	41.5	502	2	C96960	uncharacterized co
424	34	41.5	256	1	BVECBH	biotin biosynthesi	497	34	41.5	503	2	S31940	starch-degrading e
425	34	41.5	256	2	E86006	biotin biosynthesi	498	34	41.5	503	2	C97262	N-terminal domain
426	34	41.5	256	2	F91160	biotin biosynthesi	499	34	41.5	529	2	AC2112	hypothetical prote
427	34	41.5	259	2	G84685	probable DOF zinc	500	34	41.5	547	2	T00977	probable pectinest
428	34	41.5	259	2	C69122	hypothetical prote	501	34	41.5	554	2	H86772	alpha-acetolactate
429	34	41.5	261	2	B69095	cobalt transport m	502	34	41.5	563	1	A47078	phenol catabolic p
430	34	41.5	261	2	G84376	hypothetical prote	503	34	41.5	563	1	S47095	phnR protein - Pse
431	34	41.5	262	2	E96747	hypothetical prote	504	34	41.5	563	2	D84717	probable cysteinyl
432	34	41.5	262	2	R5NT2D	ribosomal protein	505	34	41.5	566	1	N1PSRP	regulatory protein
433	34	41.5	266	2	T51988	ethylene responsiv	506	34	41.5	574	2	S23530	H+-transporting tw
434	34	41.5	266	2	B82655	UPP-3-O-(R-3-hydro	507	34	41.5	628	2	AE1108	transporter homolo
435	34	41.5	268	2	S05471	embryonic abundant	508	34	41.5	628	2	AG1469	transporter homolo
436	34	41.5	274	1	R5NT2	ribosomal protein	509	34	41.5	635	2	F76371	hypothetical prote
437	34	41.5	274	2	S78397	ribosomal protein	510	34	41.5	652	2	F90561	hypothetical prote
438	34	41.5	286	1	R5SP2	ribosomal protein	511	34	41.5	673	1	C906C	collagen alpha 1(I
439	34	41.5	289	2	AC0794	conserved hypotet	512	34	41.5	674	2	T05264	probable serine/th
440	34	41.5	301	2	AC0295	conserved hypotet	513	34	41.5	681	2	H83044	2,4-dienoyl-CoA re
441	34	41.5	301	2	S77169	hypothetical prote	514	34	41.5	701	2	E70827	probable fuaA prot
442	34	41.5	302	1	PAFF1A	phosphoprotein pho	515	34	41.5	712	2	T48961	hypothetical prote
443	34	41.5	302	2	S29396	phosphoprotein pho	516	34	41.5	729	2	F83725	hypothetical prote
444	34	41.5	302	2	AE2520	hypothetical prote	517	34	41.5	742	2	J90326	alcohol dehydrogen
445	34	41.5	303	2	B43451	Na+/K+-exchanging	518	34	41.5	749	2	A95068	hypothetical prote
446	34	41.5	305	2	S44767	C29E4.1 protein -	519	34	41.5	751	2	A02168	phycochrome-like p
447	34	41.5	311	2	T00087	ramnosyltransfera	520	34	41.5	751	2	B97936	5-methyltetrahydro
448	34	41.5	312	2	C87562	conserved hypotet	521	34	41.5	797	2	F79361	probable transcrip
449	34	41.5	313	2	AC0622	probable bacteriop	522	34	41.5	863	2	JC7537	beta-N-acetylgluco
450	34	41.5	313	2	AH0734	probable bacteriop	523	34	41.5	863	2	C38252	granulocyte colony
451	34	41.5	316	2	S25188	transcription acti	524	34	41.5	864	2	JH0438	penicillin-binding
452	34	41.5	321	2	E75540	hypothetical prote	525	34	41.5	903	2	T44153	hypothetical prote
453	34	41.5	335	2	H86363	F19G10.7 protein -	526	34	41.5	905	2	S75035	H+-exporting ATPas
454	34	41.5	337	2	G83200	probable oxidoredu	527	34	41.5	908	2	AC0967	Magnesium transport
455	34	41.5	340	2	S11143	class I histocampa	528	34	41.5	908	2	B39083	Mg2+-transporting
456	34	41.5	341	2	AI1039	phage integrase [i	529	34	41.5	976	2	G82209	GGDEF family prote
457	34	41.5	342	2	QJ1651	SSL2 protein - hum	530	34	41.5	1021	2	F88568	protein F40P12.5 [
458	34	41.5	344	2	T39023	hypothetical prote	531	34	41.5	1024	2	G71434	probable limonene
459	34	41.5	345	2	C71046	hypothetical prote	532	34	41.5	1031	2	B97372	DNA polymerase I (
460	34	41.5	349	2	A75169	probable iron (III	533	34	41.5	1031	2	A12589	DNA polymerase I (
461	34	41.5	352	2	D81718	RNA (Guanine-N1) -	534	34	41.5	1033	2	S54506	probable membrane
462	34	41.5	358	2	T52073	ER-associated Hsp4	535	34	41.5	1050	2	T27753	hypothetical prote
463	34	41.5	365	2	S10847	collagen alpha 2(I	536	34	41.5	1099	2	T18713	hypothetical prote
464	34	41.5	374	2	T46065	hypothetical prote	537	34	41.5	1120	2	T14275	myosin-like protei
465	34	41.5	374	2	T43966	hypothetical prote	538	34	41.5	1130	2	A48843	MHC class II trans
466	34	41.5	376	2	S45763	hypothetical prote	539	34	41.5	1153	2	A97179	ATP-dependent exon
467	34	41.5	377	2	AE0589	galactosyltransfer	540	34	41.5	1191	2	S27329	DNA topoisomerase

541	34	41.5	1192	1	ISXNAS	DNA topoisomerase	614	33	40.2	242	2	G83948	3-oxoacyl-[acyl ca
542	34	41.5	1194	2	E96624	hypothetical prote	615	33	40.2	244	2	A96806	hypothetical prote
543	34	41.5	1263	2	T19472	hypothetical prote	616	33	40.2	246	2	AC2504	hypothetical prote
544	34	41.5	1337	2	T30291	dextranase - Strept	617	33	40.2	250	2	T01604	hypothetical prote
545	34	41.5	1392	2	A43336	microtubule-vesicl	618	33	40.2	254	2	D95360	probable GntR-fami
546	34	41.5	1418	2	T45467	collagen alpha 1(I)	619	33	40.2	261	2	T04301	beta-expansin - ri
547	34	41.5	1419	2	A41182	collagen alpha 1(I)	620	33	40.2	262	2	S49311	2,4-dihydroxyhept-
548	34	41.5	1427	2	S74293	SRB8 protein - yea	621	33	40.2	262	2	A95334	hypothetical prote
549	34	41.5	1427	2	S22695	restin - human	622	33	40.2	263	2	AD2517	hypothetical prote
550	34	41.5	1486	1	B40333	collagen alpha 1(I)	623	33	40.2	268	2	S14068	seed protein precu
551	34	41.5	1487	1	CGHUC	collagen alpha 1(I)	624	33	40.2	271	2	S18730	aminoglycoside N3'
552	34	41.5	1487	2	B41182	collagen alpha 1(I)	625	33	40.2	275	2	T11810	ribosomal protein
553	34	41.5	1492	2	A40333	collagen alpha 1(I)	626	33	40.2	275	2	T02334	probable urease ac
554	34	41.5	1763	2	T17465	rifamycin polyketi	627	33	40.2	275	2	T52333	probable urease p
555	34	41.5	1846	2	T10670	hypothetical prote	628	33	40.2	275	2	G90221	conserved hypotet
556	34	41.5	2254	2	T09053	low voltage-activa	629	33	40.2	275	2	A84429	probable S-lucos g
557	34	41.5	4116	2	T13719	calo protein - fru	630	33	40.2	276	2	T07531	ribosomal protein
558	34	41.5	4735	2	T17463	rifamycin polyketi	631	33	40.2	276	2	T47351	hypothetical prote
559	34	41.5	6420	2	T30283	polyketide synthas	632	33	40.2	280	2	T00718	calcium-dependent
560	34	41.5	26926	1	T38344	titin, cardiac mus	633	33	40.2	288	2	D85072	hypothetical prote
561	33.5	40.9	59	2	F69854	hypothetical prote	634	33	40.2	291	2	T20666	hypothetical prote
562	33.5	40.9	123	2	S30529	ig heavy chain V r	635	33	40.2	292	2	E70018	sugar permease hom
563	33.5	40.9	124	2	B55257	ig gamma heavy cha	636	33	40.2	293	1	DBP32	helix-destabilizin
564	33.5	40.9	205	2	H70468	phosphoribosyl-AMP	637	33	40.2	293	1	DBP36	helix-destabilizin
565	33.5	40.9	239	2	C70486	conserved hypotet	638	33	40.2	297	2	AF0568	carbamate kinase (
566	33.5	40.9	254	2	S34724	probable oxidoredu	639	33	40.2	299	2	F83211	hypothetical prote
567	33.5	40.9	284	2	T09843	amino acid transpo	640	33	40.2	299	2	A56663	capsid protein VPI
568	33.5	40.9	289	2	G97122	pseudouridine synt	641	33	40.2	300	2	T33232	hypothetical prote
569	33.5	40.9	351	2	T18066	hypothetical prote	642	33	40.2	301	1	DBP34	helix-destabilizin
570	33.5	40.9	381	2	E71194	probable N2,N2-dim	643	33	40.2	301	2	H95308	probable ABC trans
571	33.5	40.9	422	2	G72227	hypothetical prote	644	33	40.2	308	2	S11553	oligopeptide trans
572	33.5	40.9	486	2	T10100	amino acid transpo	645	33	40.2	308	2	E95220	hypothetical prote
573	33.5	40.9	503	2	E91055	hypothetical prote	646	33	40.2	308	2	T35303	hypothetical prote
574	33.5	40.9	503	2	A85900	hypothetical prote	647	33	40.2	311	2	B95222	BirA bifunctional
575	33.5	40.9	954	2	S46105	glucan 1,4-alpha-g	648	33	40.2	311	2	A99086	bioti-[acetyl-CoA-
576	33.5	40.9	1112	2	S70522	cyclic nucleotide	649	33	40.2	314	2	A82517	hypothetical prote
577	33.5	40.9	1353	2	T00347	hypothetical prote	650	33	40.2	316	2	T27194	hypothetical prote
578	33.5	40.9	1401	2	T30247	Werner syndrome pr	651	33	40.2	317	2	D87395	metallo-beta-lacta
579	33	40.2	20	2	PM0003	chlorophyll a/b-bi	652	33	40.2	323	2	E98084	hypothetical prote
580	33	40.2	61	2	B34123	depressant insect	653	33	40.2	323	2	B48067	ethanolamine-phosp
581	33	40.2	84	2	E83004	glutaredoxin PA512	654	33	40.2	325	1	A71887	probable GMP reduc
582	33	40.2	89	2	A97705	hypothetical prote	655	33	40.2	325	2	T26180	hypothetical prote
583	33	40.2	91	2	T05920	probable cysteine	656	33	40.2	325	2	T29604	hypothetical prote
584	33	40.2	97	2	AB3591	hypothetical prote	657	33	40.2	327	1	F64626	probable GMP reduc
585	33	40.2	102	2	T17738	hypothetical prote	658	33	40.2	328	2	T44931	mOCR protein [impo
586	33	40.2	103	2	T10920	3C3.11 protein - S	659	33	40.2	329	2	F72413	conserved hypotet
587	33	40.2	105	2	A13472	hypothetical prote	660	33	40.2	330	2	T51834	transcription fact
588	33	40.2	115	2	E97483	hypothetical prote	661	33	40.2	332	1	WMBP16	gene 16 protein -
589	33	40.2	122	2	T44906	hypothetical prote	662	33	40.2	332	1	WMBP26	gene 16 protein -
590	33	40.2	126	2	A23034	CI990 resistance p	663	33	40.2	332	2	B75011	hypothetical prote
591	33	40.2	127	2	A54670	RNA polymerase II	664	33	40.2	342	2	F85089	hypothetical prote
592	33	40.2	128	2	E72804	gp39 protein - Myc	665	33	40.2	343	2	T36590	hypothetical prote
593	33	40.2	133	2	G54445	formylmethanofuran	666	33	40.2	351	2	AB1955	S-adenosylmethioni
594	33	40.2	135	2	G87686	hypothetical prote	667	33	40.2	352	2	F82284	hypothetical prote
595	33	40.2	136	2	T34316	hypothetical prote	668	33	40.2	353	2	C85572	hypothetical prote
596	33	40.2	139	2	S76176	hypothetical prote	669	33	40.2	353	2	E90721	hypothetical prote
597	33	40.2	157	2	B72753	hypothetical prote	670	33	40.2	354	2	C82038	nitrogen regulatio
598	33	40.2	157	2	D84053	hypothetical prote	671	33	40.2	355	2	F70983	probable serine pr
599	33	40.2	159	2	AC3609	transcription regu	672	33	40.2	361	2	A86386	probable DNA-bindi
600	33	40.2	160	2	JO0542	185K secretory pro	673	33	40.2	362	2	C64807	ybG0 protein - Esc
601	33	40.2	171	2	D55853	aggA protein precu	674	33	40.2	362	2	F71872	hypothetical prote
602	33	40.2	185	2	D69976	conserved hypotet	675	33	40.2	362	2	T42689	hypothetical prote
603	33	40.2	196	2	A48150	hibernation-relate	676	33	40.2	367	2	H64696	lipopolysaccharide
604	33	40.2	203	2	G86785	acetyltransferase	677	33	40.2	367	2	H71822	lipopolysaccharide
605	33	40.2	203	2	H86834	maltose O-acetyltr	678	33	40.2	369	2	B70968	hypothetical prote
606	33	40.2	203	2	S53587	probable membrane	679	33	40.2	369	2	F95353	probable ABC trans
607	33	40.2	207	2	T17084	NADH2 dehydrogenas	680	33	40.2	373	2	T26030	hypothetical prote
608	33	40.2	208	2	T17087	NADH2 dehydrogenas	681	33	40.2	377	2	C69858	conserved hypotet
609	33	40.2	211	2	D96507	hypothetical prote	682	33	40.2	379	2	G87046	probable acyltrans
610	33	40.2	226	2	H81295	probable two-compo	683	33	40.2	384	2	G75567	conserved hypotet
611	33	40.2	230	2	D86352	protein T26F17.14	684	33	40.2	386	1	A29984	alanine racemase (
612	33	40.2	239	2	E81697	ribose 5-phosphate	685	33	40.2	386	2	T44545	contractile tail s
613	33	40.2	241	2	A53014	chloride conductan	686	33	40.2	387	2	H72299	hypothetical prote



687	33	40.2	387	2	H72273	conserved hypothet	760	33	40.2	628	2	T09785	hydroxymethylgluta
688	33	40.2	389	2	G95939	probable hippurate	761	33	40.2	634	2	S31925	beta-fructofuranos
689	33	40.2	392	2	AE2016	phosphoserine amin	762	33	40.2	636	1	S31157	beta-fructofuranos
690	33	40.2	394	2	S62009	probable membrane	763	33	40.2	636	1	S31155	beta-fructofuranos
691	33	40.2	394	2	AD0842	probable transmemb	764	33	40.2	639	2	A56126	peroxisomal target
692	33	40.2	395	2	T33677	hypothetical prote	765	33	40.2	640	1	A86657	fructose-bisphosph
693	33	40.2	398	2	A95870	hypothetical prote	766	33	40.2	640	2	T09534	probable beta-fruc
694	33	40.2	399	2	D83535	aromatic amino aci	767	33	40.2	640	2	AB2251	glucose inhibited
695	33	40.2	403	1	A10699	probable pathogeni	768	33	40.2	644	2	D83971	stage V sporulatio
696	33	40.2	404	1	S62440	mevalonate kinase	769	33	40.2	656	2	G85731	Rhs element associ
697	33	40.2	408	2	B71047	probable ferredoxi	770	33	40.2	661	2	S37591	beta-fructofuranos
698	33	40.2	408	2	C86903	hypothetical prote	771	33	40.2	666	2	T08904	probable long-chai
699	33	40.2	408	2	AC1840	hypothetical prote	772	33	40.2	667	2	T07929	probable long-chai
700	33	40.2	409	2	T35598	hypothetical prote	773	33	40.2	681	2	D84019	methylmalonyl-CoA
701	33	40.2	416	2	D83386	hypothetical prote	774	33	40.2	686	2	JC5708	villin-like protei
702	33	40.2	420	1	DCECD	diaminopimelate de	775	33	40.2	694	2	C72761	hypothetical prote
703	33	40.2	420	2	AG0867	diaminopimelate de	776	33	40.2	698	2	E85369	hypothetical prote
704	33	40.2	420	2	B85936	diaminopimelate de	777	33	40.2	698	2	T10682	hypothetical prote
705	33	40.2	420	2	G91090	diaminopimelate de	778	33	40.2	698	2	AD2985	oxidoreductase Atu
706	33	40.2	429	2	C75013	phosphoribosylglyc	779	33	40.2	706	2	F87683	peptidase M13 fami
707	33	40.2	433	2	C71138	phosphoribosylglyc	780	33	40.2	708	2	S52317	guinohemoprotein e
708	33	40.2	437	2	S73284	hypothetical prote	781	33	40.2	714	2	T16126	hypothetical prote
709	33	40.2	437	2	G82777	glutamate symport	782	33	40.2	728	2	H82965	DNA helicase II PA
710	33	40.2	439	2	S16530	xylose isomerase (	783	33	40.2	729	2	C98298	probable oxidoredu
711	33	40.2	445	2	G81200	UDP-N-acetylmuramo	784	33	40.2	735	2	A69146	hypothetical prote
712	33	40.2	445	2	D81777	UDP-N-acetylmuramo	785	33	40.2	737	2	AE1306	heavy metal-transp
713	33	40.2	452	2	T28094	hypothetical prote	786	33	40.2	737	2	AE1678	heavy metal-transp
714	33	40.2	455	2	T23712	hypothetical prote	787	33	40.2	737	2	C71122	hypothetical prote
715	33	40.2	458	2	A90565	hypothetical prote	788	33	40.2	759	2	S76989	hypothetical prote
716	33	40.2	464	2	T28818	hypothetical prote	789	33	40.2	775	2	A61228	lectin-like adhesi
717	33	40.2	466	2	AH3100	amidohydrolase lim	790	33	40.2	786	2	A35466	hypothetical prote
718	33	40.2	466	2	B98186	probable hydrolase	791	33	40.2	797	2	B95377	progestrone recep
719	33	40.2	467	2	I56896	gene gfi-2 protein	792	33	40.2	807	2	T24110	hypothetical prote
720	33	40.2	469	2	JO1644	exo-alpha-sialidas	793	33	40.2	808	1	OPKEX	glucose dehydrogen
721	33	40.2	471	2	S08333	transcription fact	794	33	40.2	819	2	A47018	hypothetical prote
722	33	40.2	471	2	T50016	hypothetical prote	795	33	40.2	819	2	B87386	hypothetical prote
723	33	40.2	474	2	T03126	hypothetical prote	796	33	40.2	835	2	JP0076	epidermal growth f
724	33	40.2	475	2	E90318	medium-chain-fatty	797	33	40.2	843	2	A27131	Mg2+-transporting
725	33	40.2	477	2	E86249	hypothetical prote	798	33	40.2	872	2	S73948	alanine-tRNA ligas
726	33	40.2	480	2	S34547	H+-transporting tw	799	33	40.2	878	2	B84977	Mg(2+) transport A
727	33	40.2	485	2	A84859	probable cytochrom	800	33	40.2	899	2	AG0202	hypothetical prote
728	33	40.2	489	2	S63401	hypothetical prote	801	33	40.2	942	2	C96574	hypothetical prote
729	33	40.2	490	2	E83202	outer membrane pro	802	33	40.2	944	2	AC2073	two-component sens
730	33	40.2	490	2	JS0586	algG protein precu	803	33	40.2	959	2	T00246	DNA polymerase V -
731	33	40.2	490	2	T41039	probable transcrip	804	33	40.2	974	2	T14076	probable villin [i
732	33	40.2	492	2	AH2809	hypothetical prote	805	33	40.2	984	2	AE0290	insecticidal toxin
733	33	40.2	495	2	B35721	nicotinic acetylch	806	33	40.2	993	2	A38437	probable homeotic
734	33	40.2	499	2	AD2262	amidophosphoribosy	807	33	40.2	1008	2	H72310	conserved hypothet
735	33	40.2	507	2	B83988	proline transport	808	33	40.2	1017	2	T08553	hypothetical prote
736	33	40.2	509	2	T06226	probable beta-fruc	809	33	40.2	1019	2	T11560	pol polyprotein -
737	33	40.2	509	2	H87389	conserved hypothet	810	33	40.2	1025	1	JC1266	beta-galactosidase
738	33	40.2	516	2	AH2417	hypothetical prote	811	33	40.2	1029	2	T05050	protein kinase hom
739	33	40.2	517	2	A30992	probable nicotinic	812	33	40.2	1036	1	GNLJG2	HIV-1 retropepsin
740	33	40.2	520	2	S45702	leukotriene-B4 20-	813	33	40.2	1039	2	S46347	pol polyprotein -
741	33	40.2	520	2	S45753	probable membrane	814	33	40.2	1046	2	T42720	cytoplasmic linker
742	33	40.2	523	2	C97588	ribose ABC transpo	815	33	40.2	1050	2	G86582	exodeoxyribonuclea
743	33	40.2	536	2	A36395	spore wall maturat	816	33	40.2	1050	2	H72041	exodeoxyribonuclea
744	33	40.2	544	2	T07593	pectinesterase (EC	817	33	40.2	1054	2	C81624	HIV-1 retropepsin
745	33	40.2	548	2	C82698	electron transfer	818	33	40.2	1054	1	GNLJG5	HIV-1 retropepsin
746	33	40.2	555	2	D90369	NADH oxidase SSO20	819	33	40.2	1056	1	GNLJG3	pol polyprotein -
747	33	40.2	555	2	C83444	probable AMP-bindi	820	33	40.2	1058	2	S08436	hypothetical prote
748	33	40.2	556	2	A44441	B-cell antigen CD1	821	33	40.2	1070	2	T34385	replication factor
749	33	40.2	557	4	EBRTMS	IGF-binding protei	822	33	40.2	1092	2	T18305	replication factor
750	33	40.2	563	1	SL1175	choleline transport	823	33	40.2	1092	2	T18306	probable histidine
751	33	40.2	569	2	A46462	r cell activation	824	33	40.2	1122	2	T00441	DNA-directed DNA p
752	33	40.2	575	2	C83313	probable type II s	825	33	40.2	1139	2	D64503	protein R06B10.1 l
753	33	40.2	576	2	T40476	hypothetical prote	826	33	40.2	1236	2	A88392	cryptic nitrate re
754	33	40.2	586	2	B90659	Rhs core protein [	827	33	40.2	1246	2	G90887	cryptic nitrate re
755	33	40.2	586	2	H85509	hypothetical prote	828	33	40.2	1246	2	B85730	nitrate reductase
756	33	40.2	600	2	H70448	G-protein LepA - A	829	33	40.2	1246	2	G84899	respiratory nitrat
757	33	40.2	607	2	A95122	TnS252, relaxase [	830	33	40.2	1247	2	AH0648	hypothetical prote
758	33	40.2	608	2	T06632	hypothetical prote	831	33	40.2	1262	2	T25168	epidermal growth f
759	33	40.2	612	2	T08602	protein tipD - sli	832	33	40.2	1330	1	GOPFE	

833	33	40.2	1374	2	D85390	myosin-like protei	906	32	39.0	96	2	B41979	neuropeptide Y pre
834	33	40.2	1375	2	T05200	myosin heavy chain	907	32	39.0	98	2	PH1147	Ig heavy chain V r
835	33	40.2	T13822	2	T13822	frazzled gene prot	908	32	39.0	99	2	D38601	Ig kappa chain V r
836	33	40.2	A29637	2	A29637	position-specific	909	32	39.0	100	2	B70977	hypothetical prote
837	33	40.2	H90698	2	H90698	RhaD core protein	910	32	39.0	101	1	GMHUB	gastirin precursor
838	33	40.2	H85549	2	H85549	hypothetical prote	911	32	39.0	105	2	A82456	probable acetyltra
839	33	40.2	E30886	2	E30886	RhaG core protein	912	32	39.0	107	2	B81964	conserved hypothet
840	33	40.2	E85509	2	E85509	hypothetical prote	913	32	39.0	107	2	C81020	cyay protein NMA19
841	33	40.2	E90658	2	E90658	RhaG core protein	914	32	39.0	114	2	A27563	Ig heavy chain V r
842	33	40.2	T27969	2	T27969	hypothetical prote	915	32	39.0	114	2	T03081	hypothetical prote
843	33	40.2	H64780	2	H64780	RhaD protein precu	916	32	39.0	121	2	C85515	unknown protein en
844	33	40.2	A70982	2	A70982	probable ATP-depen	917	32	39.0	115	2	B85573	succinate dehydrog
845	33	40.2	1513	2	1513	frazzled gene prot	918	32	39.0	115	2	C90722	succinate dehydrog
846	33	40.2	1526	2	1526	glutamate synthase	919	32	39.0	116	2	D86510	hypothetical prote
847	33	40.2	E82085	2	E82085	hypothetical prote	920	32	39.0	117	2	H90664	hypothetical prote
848	33	40.2	1549	2	1549	peptidoglycan boun	921	32	39.0	121	2	G95319	hypothetical prote
849	33	40.2	1660	2	1660	hypothetical glyci	922	32	39.0	121	2	C85515	unknown protein en
850	33	40.2	1700	2	1700	Balbani ring 3 pr	923	32	39.0	128	2	AC0264	probable pyrophosp
851	33	40.2	1712	1	CGHU2B	collagen alpha 2(I	924	32	39.0	130	2	F69044	mutator MutT prote
852	33	40.2	1797	2	F59195	cell surface glyco	925	32	39.0	133	2	F86816	hypothetical prote
853	33	40.2	1871	2	S27938	hypothetical prote	926	32	39.0	134	2	F87352	mutator mutT prote
854	33	40.2	1906	1	S68235	myosin-light-chain	927	32	39.0	141	2	C86521	hypothetical prote
855	33	40.2	2212	2	A41098	calcium channel pr	928	32	39.0	141	2	G72100	hypothetical prote
856	33	40.2	2245	2	T18278	myosin heavy chain	929	32	39.0	141	2	C82466	conserved hypothet
857	33	40.2	2273	2	146477	calcium channel BI	930	32	39.0	142	1	R3IT8	ribosomal protein
858	33	40.2	2298	2	T49648	hypothetical prote	931	32	39.0	144	2	T00763	hypothetical prote
859	33	40.2	2348	2	AD1841	hypothetical prote	932	32	39.0	147	2	CO8864	conserved hypothet
860	33	40.2	2424	2	146480	calcium channel BI	933	32	39.0	154	1	A55692	anaerobic ribonucl
861	33	40.2	2761	2	T21064	hypothetical prote	934	32	39.0	154	2	AB1058	anaerobic ribonucl
862	33	40.2	2831	2	T31419	cyclic beta 1-2 gl	935	32	39.0	154	2	F91280	anaerobic ribonucl
863	33	40.2	2924	2	T18378	variant-specific 8	936	32	39.0	154	2	F86121	anaerobic ribonucl
864	33	40.2	3029	2	S76109	hypothetical prote	937	32	39.0	154	2	H83160	hypothetical prote
865	33	40.2	5005	2	F82884	hypothetical prote	938	32	39.0	158	2	T27757	hypothetical prote
866	33	40.2	5069	2	T17464	rifamycin polyketi	939	32	39.0	159	2	AD3506	sensor protein chv
867	33	40.2	10223	2	T30225	polyketide synthas	940	32	39.0	159	2	F71046	hypothetical prote
868	32.5	39.6	101	2	S13692	Ig heavy chain V r	941	32	39.0	160	2	F86186	YUP812.17 import
869	32.5	39.6	111	2	S13693	Ig heavy chain V r	942	32	39.0	164	2	B90398	mutT-like protein
870	32.5	39.6	112	2	S13690	Ig heavy chain V r	943	32	39.0	167	2	H90412	hypothetical prote
871	32.5	39.6	115	2	S13694	Ig heavy chain V r	944	32	39.0	167	2	T48744	hypothetical prote
872	32.5	39.6	116	2	S13691	Ig heavy chain V r	945	32	39.0	172	2	C75295	MutT/nudix family
873	32.5	39.6	118	2	C30560	Ig heavy chain V r	946	32	39.0	172	2	C83269	hypothetical prote
874	32.5	39.6	119	2	FL0085	Ig heavy chain V r	947	32	39.0	174	2	S67663	hypothetical prote
875	32.5	39.6	124	2	S06824	Ig heavy chain V r	948	32	39.0	175	2	S36749	transcription fact
876	32.5	39.6	137	2	S03326	Ig heavy chain pre	949	32	39.0	175	2	AH3251	conserved hypothet
877	32.5	39.6	182	2	T49852	hypothetical prote	950	32	39.0	178	2	G87307	hypothetical prote
878	32.5	39.6	194	2	D75428	hypothetical prote	951	32	39.0	179	1	A42840	ribosomal protein
879	32.5	39.6	223	2	B71514	hypothetical prote	952	32	39.0	179	2	JQ0345	ubiquinol-cytochro
880	32.5	39.6	242	2	A57504	ycli protein impor	953	32	39.0	179	2	T49508	hypothetical prote
881	32.5	39.6	242	2	AD2722	conserved hypothet	954	32	39.0	180	2	C81557	KDO-transferrase 2
882	32.5	39.6	285	2	A84953	pantoate-beta-alan	955	32	39.0	185	2	SL2205	hypothetical prote
883	32.5	39.6	288	2	D75286	serine proteinase	956	32	39.0	188	2	A13391	hypothetical cytos
884	32.5	39.6	319	2	H83253	conserved hypothet	957	32	39.0	196	2	T35037	hypothetical prote
885	32.5	39.6	346	2	S46993	elk ligand - human	958	32	39.0	207	2	D90526	conserved hypothet
886	32.5	39.6	364	2	T17584	probable beta-1,3-	959	32	39.0	209	2	JC7239	peroxiredoxin V -
887	32.5	39.6	391	2	D64366	hypothetical prote	960	32	39.0	213	2	JE0298	transcription fact
888	32.5	39.6	433	2	C69222	hypothetical prote	961	32	39.0	214	2	E82679	chaperone XF1452 [
889	32.5	39.6	436	2	JC1497	alpha-amino-epsilo	962	32	39.0	216	2	JE0297	DRE/CRT-binding pr
890	32.5	39.6	446	2	H81208	conserved hypothet	963	32	39.0	216	2	JE0299	transcription facto
891	32.5	39.6	446	2	C82000	probable integral	964	32	39.0	216	2	T51830	transcription fact
892	32.5	39.6	449	2	T12495	hypothetical prote	965	32	39.0	216	2	F70073	hypothetical prote
893	32.5	39.6	594	2	B82994	probable secretion	966	32	39.0	220	2	AB0084	probable DedA-fami
894	32.5	39.6	615	2	A05269	collagen alpha 1(I	967	32	39.0	226	2	T00409	ethylene-responsiv
895	32.5	39.6	623	2	T18892	hypothetical prote	968	32	39.0	226	2	T21380	hypothetical prote
896	32.5	39.6	772	2	S62481	hypothetical prote	969	32	39.0	227	2	T05800	probable transcrip
897	32.5	39.6	940	2	T00056	hypothetical prote	970	32	39.0	228	2	A38260	heat shock protein
898	32.5	39.6	1055	2	T05663	hypothetical prote	971	32	39.0	233	2	E87550	hypothetical prote
899	32.5	39.6	1154	2	A13431	chromosome segrega	972	32	39.0	235	2	S43513	hemoglobin linker
900	32.5	39.6	1557	2	T02859	probable serine/th	973	32	39.0	237	2	G75261	cytidylate kinase
901	32	39.0	40	2	S71917	hemoglobin, extrac	974	32	39.0	239	2	B83709	hypothetical prote
902	32	39.0	70	2	T08334	hypothetical prote	975	32	39.0	240	2	E98999	uridylylate kinase
903	32	39.0	86	2	S38920	hypothetical prote	976	32	39.0	241	2	B71128	probable iron (III
904	32	39.0	86	2	G70611	hypothetical prote	977	32	39.0	244	2	T02511	DREB-like AP2 doma
905	32	39.0	91	2	G84766	hypothetical prote	978	32	39.0	244	2	JC2379	cell-specific heli

979	32	39.0	245	2	S00413	photosystem II oxy
980	32	39.0	246	2	T20049	hypothetical prote
981	32	39.0	247	1	A64590	probable 3-oxoacyl
982	32	39.0	247	2	B71923	3-oxoacyl-[acyl-ca
983	32	39.0	247	2	I48149	serotonin receptor
984	32	39.0	247	2	H69374	NH(3)-dependent NA
985	32	39.0	249	2	F83313	hydroxyacylglutath
986	32	39.0	251	2	F83972	enoyl-CoA hydratase
987	32	39.0	251	2	H83447	hypothetical prote
988	32	39.0	251	2	S75084	hypothetical prote
989	32	39.0	252	2	AG2270	hypothetical prote
990	32	39.0	253	2	A45821	probable transposase
991	32	39.0	253	2	A86884	transposase of IS9
992	32	39.0	253	2	C86728	transposase of IS9
993	32	39.0	253	2	F86727	transposase of IS9
994	32	39.0	253	2	S12722	probable transposase
995	32	39.0	255	2	E72226	conserved hypothet
996	32	39.0	257	2	G72338	hypothetical prote
997	32	39.0	257	2	T01254	hypothetical prote
998	32	39.0	258	1	Q00WC1	coat protein - cas
999	32	39.0	259	1	YXBYT	thymidylate synthase
1000	32	39.0	259	2	D97859	methionyl aminopep

## ALIGNMENTS

RESULT 1  
S44953  
lmbf protein - Streptomyces lincolnensis  
C:Species: Streptomyces lincolnensis  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004  
C:Accession: S69815; S44953  
R:Peschke, U.; Schmidt, H.; Zhang, H.Z.; Piepersberg, W.  
Mol. Microbiol. 16, 1137-1156, 1995  
A:Title: Molecular characterization of the lincomycin-production gene cluster of Streptomyces lincolnensis  
A:Reference number: S69805; MUID: S6020646; PMID: 8577249  
A:Accession: S69815  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-426 <PE2>  
A:Cross-references: UNIPROT:Q54359; UNIPARC:UPI000003E0A; EMBL:X79146; NID:9499194; PID:9499194  
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, May 1994  
C:Gene: lmbf  
A:Gene: lmbf

Query Match	57.3%	Score 47;	DB 2;	Length 426;
Best Local Similarity	60.0%;	Pred. No. 5.3;		
Matches	9;	Mismatches	1;	
Conservative		Indels	5;	Gaps
				0;

Qy 1 RPKKKVWLGETSSAY 15  
db 395 RPWFKKVWLGRDSSVF 409

RESULT 2  
NCEX5  
exodeoxyribonuclease V (EC 3.1.11.5) 135K chain - *Escherichia coli* (strain K-12)  
N;Alternate names: exonuclease 135K polypeptide; recBC DNase 135K polypeptide  
C;Species: *Escherichia coli*  
C;Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 09-Jul-2004  
C;Accession: A25532; E65064  
R;Finch, P.W.; Storey, A.; Chapman, K.E.; Brown, K.; Hickson, I.D.; Emmerson, P.T.  
Nucleic Acids Res 14, 8573-8582, 1986  
A;Title: Complete nucleotide sequence of the *Escherichia coli* recB gene.  
A;Reference number: A25532; MUID:87066729; PMID:3537960  
A;Accession: A25532  
A;Molecule type: DNA  
A;Residues: 1-1180 <FIN>  
A;Cross-references: UNIPROT:P08394; UNIPARC:UPI000012A2D0; GB:X04581; NID:X042680;  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997

A>Title: The complete genome sequence of *Escherichia coli* K-12.  
A/Reference number: A64720; MUID:97426617; PMID:9278503  
A/Accession: E65064  
A/Status: preliminary; nucleic acid sequence not shown; transla  
A/Molecule type: DNA  
A/Residues: 1-1180 <BLAT>  
A/Cross-references: UNIPARC:UPI000012A2D0; GB:AE000365; GB:U0000  
A/Experimental source: strain K-12, substrain MG1655  
C/Comment: This enzyme is required for efficient DNA repair; it  
11 of these activities require concomitant hydrolysis of ATP.

C/Genetics:  
A/Gene: recB  
A/Map position: 61 min  
C/Superfamily: exodeoxyribonuclease V 135K chain  
C/Keywords: ATP; DNA repair; hydrolase; nucleotide binding; P-loop  
F:23-30/Region: nucleotide-binding motif A (P-loop)

Query Match	Score 46;	DB 1;	Length 1180;
Best Local Similarity	56.1%;		
Pred. No. 21;	75.0%;		
Matches 9;			
Conservative	0;	Mismatches	3: Indels 0: Gaps 0:

QY	4	KKVWLGETSSAY	15
Db	1082	KSNWLGEDSSAY	10

RESULT 3

G85933  
 DNA helicase RecB [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
 C:Accession: G85933  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.;  
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Ap-  
 Nature 409 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: G85933  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1180 <STO>  
 A:Cross-references: UNIPROT:Q9X6M9; UNIPARC:UPI00000D0B55; GB:A85005174; NID:g1250517  
 A:Experimental source: strain O157:H7, substrain EDL933

Query Match 56.1%; Score 46; DB 2; Length 1180;  
Best Local Similarity 75.0%; Pred. No. 21;  
Matches 9: Conservative 0; Mismatches 3; Indels

QY 4 KKVWLGETSSAY 15  
Db 1082 KSNWLGEDSSAY 1093

## RESULT 4

E91088  
DNA helicase RecB [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C:Accession: E91088  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome  
A:Reference number: A99629; PMID:21156231; PMID:11258796

A/reslques: 1-1180 <RIN>  
A/Cross-references: UNIPROT:P08394; UNIPARC:UPT000012A2D0; GB:X04581; NID:942680; PIDN:C  
R;Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997

```

C;Genetics:
A;Gene: ECa3677
C;Superfamily: exodeoxyribonuclease V 135K chain

Query Match      56.1%; Score 46; DB 2; Length 1180;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 KKVWLGETSSAY 15
Db      1082 KSNWLGEDSSAY 1093

RESULT 5
AG2501
hypothetical protein all1791 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AG2501
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AG2501
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1596 <KUR>
A;Cross-references: UNIPROT:Q8YKV0; UNIPARC:UPI00000CEFI7; GB:BA000020; PIDN:BA078275.1;
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all1791
A;Genome: plasmid

Query Match      55.5%; Score 45.5; DB 2; Length 1596;
Best Local Similarity 52.9%; Pred. No. 35;
Matches 9; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY      1 RPK---KKVWLGETSSA 14
      ||| :||| :
Db      240 RPKLYKQIKWLGEKSSQ 256

RESULT 6
D84828
AP2 domain transcription factor [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C;Accession: D84828
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: D84828
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-177 <STO>
A;Cross-references: UNIPROT:Q9S1Z0; UNIPARC:UPI00000A8651; GB:AE002093; NID:G4586052; PI
C;Genetics:
A;Gene: At2g40350
A;Map position: 2

Query Match      54.9%; Score 45; DB 2; Length 177;
Best Local Similarity 61.5%; Pred. No. 4.9;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      3 GKKVWLGETSSAY 15
      | :||| :|
Db      88 GAKLWLGETSSSY 100

C;Genetics:
A;Gene: ECa3677
C;Superfamily: exodeoxyribonuclease V 135K chain

Query Match      56.1%; Score 46; DB 2; Length 1180;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 KKVWLGETSSAY 15
Db      1082 KSNWLGEDSSAY 1093

RESULT 7
AD0783
PTS system, fructose-specific IIBC component [imported] - Salmonella enterica subsp. ent
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AD0783
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Mouton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AD0783
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-562 <PAR>
A;Cross-references: UNIPARC:UPI0000059BC9; GB:AL513382; PIDN:CAD02586.1; PID:G16503443; C
C;Genetics:
A;Gene: STY2439
C;Superfamily: phosphotransferase system enzyme II, fructose-specific; phosphotransferase
Query Match      54.9%; Score 45; DB 2; Length 562;
Best Local Similarity 66.7%; Pred. No. 15;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      3 GKKVWLGETSSA 14
      ||||| :
Db      59 GKKVWLGDIGRA 70

RESULT 8
E84594
AP2 domain transcription factor [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C;Accession: E84594
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: E84594
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-336 <STO>
A;Cross-references: UNIPROT:Q9SKT1; UNIPARC:UPI000009FA1A; GB:AE002093; NID:G4454460; PI
C;Genetics:
A;Gene: At2g20880
A;Map position: 2

Query Match      53.7%; Score 44; DB 2; Length 336;
Best Local Similarity 58.3%; Pred. No. 13;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      2 PGKKVWLGETSS 13
      || :||| :
Db      43 PGSDMWLGDASS 54

RESULT 9
E85855
PTS system, fructose-specific transport protein [imported] - Escherichia coli (strain O15
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: E85855
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

```

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: E85855

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-563 <STO>

A;Cross-references: UNIPROT:Q8XE91; UNIPARC:UPI00000D0287; GB:AE005174; NID:g12516493; E

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: fruA

C;Superfamily: phosphotransferase system enzyme II, fructose-specific; phosphotransferase

Query Match 53.7%; Score 44; DB 2; Length 563;

Best Local Similarity 66.7%; Pred. No. 22;

Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 GKKVWLGETSSA 14

Db 59 GKNVWLGDISRA 70

RESULT 10

C91011 fructose-specific PTS transport system protein ECS3059 [imported] - Escherichia coli (str

C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004

C;Accession: C91011

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gawawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A;Reference number: A99629; MUID:21156231; PMID:11258796

A;Accession: C91011

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-563 <HAY>

A;Cross-references: UNIPROT:Q8XE91; UNIPARC:UPI00000D0287; GB:BA000007; PIDN:BA836482.1;

A;Experimental source: strain O157:H7, substrain RIMD 0509952

C;Genetics:

A;Gene: ECS3059

C;Superfamily: phosphotransferase system enzyme II, fructose-specific; phosphotransferase

Query Match 53.7%; Score 44; DB 2; Length 563;

Best Local Similarity 66.7%; Pred. No. 22;

Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 GKKVWLGETSSA 14

Db 59 GKNVWLGDISRA 70

RESULT 11

A34962

phosphotransferase system enzyme II (EC 2.7.1.69), fructose-specific - Escherichia coli

C;Species: Escherichia coli

C;Date: 13-Jul-1990 #sequence\_revision 13-Jul-1990 #text\_change 09-Jul-2004

C;Accession: A34962; C37245; F64985

R;Prior, T.I.; Kornberg, H.L.

J. Gen. Microbiol. 134, 2757-2768, 1988

A;Title: Nucleotide sequence of fruA, the gene specifying enzyme II(fru) of the phosphoe

A;Reference number: A34962; MUID:89341690; PMID:3076173

A;Accession: A34962

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-563 <PRI>

A;Cross-references: UNIPROT:P20966; UNIPARC:UPI00001328CE; GB:M23196; NID:g340830; PIDN:

R;Orchard, L.M.D.; Kornberg, H.L.

Proc. R. Soc. Lond. B Biol. Sci. 242, 87-90, 1990

A;Title: Sequence similarities between the gene specifying 1-phosphofructokinase (fruK),

A;Reference number: A37245; MUID:91164390; PMID:1981619

A;Accession: C37245

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-74 <ORC>

A;Cross-references: UNIPARC:UPI0000178EC4; EMBL:X53948

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: F64985

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-563 <BLAT>

A;Cross-references: UNIPARC:UPI00001328CE; GB:AE000306; GB:U00096; NID:g1788489; PIDN:AA

C;Genetics:

A;Gene: fruA; ptsF

A;Map position: 47 min

C;Superfamily: phosphotransferase system enzyme II, fructose-specific; phosphotransferase

C;Keywords: phosphoprotein; phosphotransferase; transmembrane protein

Query Match 53.7%; Score 44; DB 2; Length 563;

Best Local Similarity 66.7%; Pred. No. 22;

Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 GKKVWLGETSSA 14

Db 59 GKNVWLGDISRA 70

RESULT 12

A83664

tRNA/rRNA methyltransferase BH0113 [imported] - Bacillus halodurans (strain C-125)

C;Species: Bacillus halodurans

C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004

C;Accession: A83664

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiran

Nucleic Acids Res. 28, 4317-4331, 2000

A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A;Reference number: A83650; MUID:20512582; PMID:11058132

A;Accession: A83664

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-249 <STO>

A;Cross-references: UNIPROT:Q9KGF2; UNIPARC:UPI00000D4202; GB:AP001507; GB:BA000004; NID:

A;Experimental source: strain C-125

C;Genetics:

A;Gene: BH0113

C;Superfamily: conserved hypothetical protein HI0860

Query Match 52.4%; Score 43; DB 2; Length 249;

Best Local Similarity 58.3%; Pred. No. 15;

Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RPKKKVWLGETS 12

Db 20 RPNKNIWIGEGS 31

RESULT 13

T51257

calmodulin-binding protein - Arabidopsis thaliana

N;Alternate names: protein T5M16.20

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 09-Jul-2004

C;Accession: T51257

R;Benes, V.; Wurmback, E.; Drzonek, H.; Ansoerge, W.; Meves, H.W.; Rudd, S.; Lemcke, K.; M

submitted to the Protein Sequence Database, August 2000

A;Reference number: Z25346

A;Accession: T51257

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1022 <BEN>

A;Cross-references: UNIPROT:Q9LET7; UNIPARC:UPI00000A94A4; EMBL:AL390921

A;Experimental source: cultivar Columbia; BAC clone T8M16

C;Genetics:

A:Map position: 3  
A: Introns: 99/3; 148/3; 412/2; 480/3; 550/1; 713/3; 856/1; 883/1; 946/3; 1006/2  
A: Note: T8M16\_20

Query Match 52.4%; Score 43; DB 2; Length 1022;  
Best Local Similarity 70.0%; Pred. No. 59;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 GKKVWLGETS 12  
| | | | |  
Db 77 GSKIWLSETS 86  
| | | | |

RESULT 14  
T50928  
calmodulin-binding protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 09-Jul-2004  
C:Accession: T50927; T50928  
R:Buaboocha, T.; Zielinski, R.E.  
submitted to the EMBL Data Library, December 1999  
A:Description: Isolation of cDNA and genomic DNA clones encoding a calmodulin-binding protein  
A:Reference number: 225280  
A:Accession: T50927  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1022 <BUA>  
A:CROSS-references: UNIPROT:Q9LDS8; UNIPARC:UPI00000A4F8A; EMBL:AF217546; PTDN:AAF28347.  
A:Accession: T50928  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1022 <BU2>  
A:CROSS-references: UNIPARC:UPI00000A4F8A; EMBL:AF217547; PTDN:AAF28348.1  
C:Genetics:  
A:Introns: 99/3; 148/3; 412/1; 480/3; 550/1; 713/3; 856/1; 883/1; 946/3; 1006/2

Query Match 52.4%; Score 43; DB 2; Length 1022;  
Best Local Similarity 70.0%; Pred. No. 59;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 GKKVWLGETS 12  
| | | | |  
Db 77 GSKIWLSETS 86  
| | | | |

RESULT 15  
AB0865  
exonuclease V chain [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AB0865  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AB0865  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1181 <PAR>  
A:CROSS-references: UNIPARC:UPI000005A3D8; GB:AL513382; PTDN:CAD02818.1; PID:g16504072;  
C:Genetics:  
A:Gene: recB  
C:Superfamily: exodeoxyribonuclease V 135K chain

Query Match 52.4%; Score 43; DB 2; Length 1181;  
Best Local Similarity 66.7%; Pred. No. 67;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 KKVWLGETSSAY 15

Db 1082 KSNWLGEDSAAY 1093  
| | | | |

RESULT 16  
C84828  
AP2 domain transcription factor [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: C84828  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: AB4420; MUID:20083487; PMID:10617197  
A:Accession: C84828  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-283 <STO>  
A:CROSS-references: UNIPROT:Q8LFR2; UNIPARC:UPI00000A3241; GB:AE002093; NID:94586051; PFI  
C:Genetics:  
A:Gene: At2g40340  
A:Map position: 2

Query Match 51.2%; Score 42; DB 2; Length 283;  
Best Local Similarity 53.8%; Pred. No. 24;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 GKKVWLGETSSAY 15  
| | | | |  
Db 35 GARLWLTFTSSY 47  
| | | | |

RESULT 17  
CB3034  
probable oxidoreductase PA4889 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 05-Oct-2004  
C:Accession: CB3034  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: AB2950; MUID:20437337; PMID:10984043  
A:Accession: CB3034  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-366 <STO>  
A:CROSS-references: UNIPROT:Q9HUS4; UNIPARC:UPI00000C5E5F; GB:AE004902; GB:AE004091; NID:  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA4889  
C:Superfamily: phthalate dioxygenase reductase

Query Match 51.2%; Score 42; DB 2; Length 366;  
Best Local Similarity 66.7%; Pred. No. 31;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 PGKVKWLGE 10  
| | | | |  
Db 186 FGQRAWLGE 194  
| | | | |

RESULT 18  
E70320  
polyribonucleotide nucleotidyltransferase - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004  
C:Accession: E70320  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Over

Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: E70320  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-775 <AQF>  
A:Cross-references: UNIPROT:O66593; UNIPARC:UPI00000562CC; GB:AE000679; NID:g2982936; PIR:AE000679  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: ptpA  
C:Superfamily: polyribonucleotide nucleotidyltransferase

Query Match 51.2%; Score 42; DB 2; Length 775;  
Best Local Similarity 53.8%; Pred. No. 65;  
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 GKKVWGETSSAY 15  
| | | | |  
DB 585 GVRVWVGEQGVY 597

RESULT 19

CGH2S  
collagen alpha 2(I) chain precursor - human  
N:Alternate names: procollagen alpha 2(I) chain  
C:Species: Homo sapiens (man)  
C:Date: 30-Jun-1989 #sequence revision 25-Aug-1995 #text change 31-Dec-2004  
C:Accession: A28500; S00824; S09176; I55311; A28472; A42165; A34405; A90567; I55311; A02865  
R:De Wet, W.; Bernard, M.; Benson-Chanda, V.; Chu, M.L.; Dickson, L.; Weil, D.; Ramirez, J. Biol. Chem. 262, 16032-16036, 1987  
A:Title: Organization of the human pro-alpha-2(I) collagen gene.  
A:Reference number: A28500; MUID:88058962; PMID:2824475  
A:Accession: A28500  
A:Molecule type: DNA; mRNA  
A:Residues: 1-248, 'N', 250-1366 <DEM>  
A:Cross-references: UNIPROT:P08123; UNIPROT:Q14038; UNIPROT:Q9UM83; UNIPROT:Q9UM66; UNIPROT:Q9UM66  
R:Kuivaniemi, H.; Tromp, G.; Chu, M.L.; Prockop, D.J. Biochem. J. 252, 633-640, 1988  
A:Title: Structure of a full-length cDNA clone for the prepro-alpha-2(I) chain of human procollagen.  
A:Reference number: S00824; MUID:88339824; PMID:3421913  
A:Accession: S00824  
A:Molecule type: mRNA  
A:Residues: 1-275, 'A', 277-332, 'V', 334-337, 'A', 339-482, 'A', 484-548, 'D', 550-765 <KU11>  
A:Cross-references: UNIPARC:UPI000016A6FC; EMBL:Y00724; NID:G30022; PIDN:CAA68709.1; PIDN:CAA68709.1  
R:Dickson, L.A.; de Wet, W.; di Liberto, M.; Weil, D.; Ramirez, F. Nucleic Acids Res. 13, 3427-3438, 1985  
A:Title: Analysis of the promoter region and the N-propeptide domain of the human procollagen alpha 2(I) chain.  
A:Reference number: S09176; MUID:85242047; PMID:4011429  
A:Accession: S09176  
A:Molecule type: DNA  
A:Residues: 1-23, '33-58, 'P', 60-93 <DIC>  
A:Cross-references: UNIPARC:UPI0000173B96; UNIPARC:UPI0000173B97; EMBL:X02488; NID:g3009  
R:Weil, D.; D'Alessio, M.; Ramirez, F.; Eyre, D.R. J. Biol. Chem. 265, 16007-16011, 1990  
A:Title: Structural and functional characterization of a splicing mutation in the pro-alpha 2(I) chain of human procollagen.  
A:Reference number: I55311; MUID:90368825; PMID:2394758  
A:Accession: I55311  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 76-93 <WE11>  
A:Cross-references: UNIPARC:UPI000006F17F; GB:M35391; NID:g189684; PIDN:AAA60041.1; PIDN:AAA60041.1  
A:Accession: A58111  
A:Molecule type: protein  
A:Residues: 23-75, 94-96 <WE12>  
A:Cross-references: UNIPARC:UPI0000173B98  
A:Note: mutant sequence from a patient with Ehlers-Danlos syndrome type VII  
R:Wirtz, M.K.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollister, D.W. J. Biol. Chem. 262, 16376-16385, 1987  
A:Title: Ehlers-Danlos syndrome type VII. Deletion of 18 amino acids comprising the N-terminal propeptide.  
A:Reference number: A28472; MUID:88059013; PMID:3680255  
A:Accession: A28472

A:Molecule type: protein  
A:Residues: 32-75, 94-111 <WIR>  
A:Cross-references: UNIPARC:UPI0000173B99  
A:Note: mutant sequence of patient with Ehlers-Danlos syndrome type VII  
R:Chiodo, A.A.; Hockey, A.; Cole, W.G. J. Biol. Chem. 267, 6361-6369, 1992  
A:Title: A base substitution at the splice acceptor site of intron 5 of the COL1A2 gene results in a mutant procollagen type VII.  
A:Reference number: A42165; MUID:92210617; PMID:1556139  
A:Accession: A42165  
A:Molecule type: mRNA  
A:Residues: 50-126 <CHI>  
A:Cross-references: UNIPARC:UPI0000173B9A  
A:Note: parts of this sequence were determined by protein sequencing; a mutant sequence [J. Biol. Chem. 264, 16804-16809, 1989]  
J. Biol. Chem. 264, 16804-16809, 1989  
A:Title: Temperature-dependent expression of a collagen splicing defect in the fibroblast culture.  
A:Reference number: A34405; MUID:89380311; PMID:2777808  
A:Accession: A34405  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 58-108 <WE13>  
A:Cross-references: UNIPARC:UPI0000173B9B; GB:J05049  
A:Note: the accession cited by the authors is not found in GenBank  
A:Note: parts of this sequence were determined by protein sequencing; a mutant having 93-108 amino acids.  
R:Click, E.M.; Bornstein, P. Biochemistry 9, 4699-4706, 1970  
A:Title: Isolation and characterization of the cyanogen bromide peptides from the alpha 1(I) chain of human procollagen.  
A:Reference number: A90567; MUID:71038625; PMID:5529814  
A:Accession: A90567  
A:Molecule type: protein  
A:Residues: 'Z', '81', 'B', '83-96; 417-447 <CLI>  
A:Cross-references: UNIPARC:UPI0000173B9C; UNIPARC:UPI0000173B9D  
A:Note: the compositions of peptides CNB1, CNB2, and CNB3 were determined; evidence for the presence of a single base mutation in intron leads to abnormal splicing of mRNA  
R:Kuivaniemi, H.; Sabol, C.; Tromp, G.; Sippola-Thiele, M.; Prockop, D.J. J. Biol. Chem. 263, 11407-11413, 1988  
A:Title: A 19-base pair deletion in the pro-alpha 2(I) gene of type I procollagen that causes an asymptomatic mother.  
A:Reference number: I55264; MUID:88298792; PMID:3403536  
A:Accession: I55264  
A>Status: translation not shown; translated from GB/EMBL/DBJ  
A:Molecule type: DNA; mRNA  
A:Residues: 145-197 <KU12>  
A:Cross-references: UNIPARC:UPI000016A6F4; GB:M21671; NID:g189521; PIDN:AAA59994.1; PIDN:AAA59994.1  
A:Note: single base mutation in intron leads to abnormal splicing of mRNA  
R:Chipman, S.D.; Shapiro, J.R.; McKinstry, M.B.; Stover, M.L.; Branson, P.; Rowe, D.W. J. Bone Miner. Res. 7, 793-805, 1992  
A:Title: Expression of mutant alpha 1(I)-procollagen in osteoblast and fibroblast cultures.  
A:Reference number: I55485; MUID:92351816; PMID:1642148  
A:Accession: I55485  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 163-181, 200-213 <CH2>  
A:Cross-references: UNIPARC:UPI0000071E28; GB:S41099; NID:g252702; PIDN:AAB22761.1; PIDN:AAB22761.1  
A:Note: mutant sequence from a patient with osteogenesis imperfecta type IV  
R:Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W. J. Biol. Chem. 245, 5042-5048, 1970  
A:Title: Comparative study of glycopeptides derived from selected vertebrate collagens. I. Amino acid composition and structure of the N-terminal propeptide.  
A:Reference number: A92069; MUID:71001508; PMID:4319110  
A:Accession: A92069  
A:Molecule type: protein  
A:Residues: 175-180 <MOR>  
A:Cross-references: UNIPARC:UPI0000014DF12  
A:Experimental source: skin  
A:Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine  
R:Fietzek, P.P.; Furthmayr, H.; Kuehn, K. Eur. J. Biochem. 47, 257-261, 1974  
A:Title: Comparative sequence studies on alpha2-CB2 from calf, human, rabbit and pig-skin.  
A:Reference number: A91224; MUID:75008198; PMID:4412529  
A:Accession: A91224  
A:Molecule type: protein  
A:Residues: 418-447 <PIE>  
A:Cross-references: UNIPARC:UPI00000173B9E



R;Tromp, G.; Prockop, D.J.  
Proc. Natl. Acad. Sci. U.S.A. 85, 5254-5258, 1988

A;Title: Single base mutation in the pro alpha 2(I) collagen gene that causes efficient  
A;Reference number: I59125; MUID:88276936; PMID:2839839

A;Accession: I59125

A;Status: translation not shown; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 520-573 <TRO>

A;Cross-references: UNIPARC:UPI0000073009; GB:M21353; NID:g180881; PIDN:AAA52053.1; PID:  
A;Note: single base mutation in intron leads to splicing out of exon 28

R;Bernard, M.P.; Myers, J.C.; Chu, M.L.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.  
Biochemistry 22, 1139-1145, 1983

A;Title: Structure of a cDNA for the proalpha-2 chain of human type I procollagen. Compa

A;Reference number: S09174; MUID:83178919; PMID:6687691

A;Accession: S09174

A;Molecule type: mRNA

A;Residues: 623-742, 'A', 744-764, 'X', 766-827, 'A', 829-830, 'P', 832-836, 'P', 838-1097, 'L', 109

A;Cross-references: UNIPARC:UPI0000173B9F; GB:J00115; GB:V00503; NID:g30123; PIDN:CAA37

A;Experimental source: skin fibroblast cells

R;Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes

Hum. Mol. Genet. 3, 2201-2206, 1994

A;Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the c

A;Reference number: I54365; MUID:95187161; PMID:7881420

A;Accession: I54365

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: mRNA

A;Residues: 663-675, 'V', 677, 'P', 679-742, 'A', 744-746 <FOR>

A;Cross-references: UNIPARC:UPI000006E0A7; GB:L47668; NID:g1009095; PIDN:AAB59577.1; PID

R;Niyibizi, C.; Bonadio, J.; Byers, P.H.; Eyre, D.R.  
J. Biol. Chem. 267, 23108-23112, 1992

A;Title: Incorporation of type I collagen molecules that contain a mutant alpha 2(I) cha

A;Reference number: I55369; MUID:93054637; PMID:1385413

A;Accession: I55369

A;Status: translated from GB/EMBL/DBDJ

A;Molecule type: mRNA

A;Residues: 665-666, 'D', 668-670 <NIY>

A;Cross-references: UNIPARC:UPI000011E7D9; GB:L00613; NID:g180888; PIDN:AAB59384.1; PID:

A;Note: mutant sequence from a patient with osteogenesis imperfecta

R;Bateman, J.F.; Hannagan, M.; Chan, D.; Cole, W.G.  
Biochem. J. 276, 765-770, 1991

A;Title: Characterization of a type I collagen alpha 2(I) glycine-586 to valine substitut

e method.

A;Reference number: A56799; MUID:91291136; PMID:2064612

A;Accession: A56799

A;Molecule type: mRNA

A;Residues: 672-675, 'V', 677, 'P', 679-681 <BAT>

A;Cross-references: UNIPARC:UPI000018B384; GB:S39878; NID:g1679911; PIDN:AAB19314.1; PID

A;Note: sequence extracted from NCBI backbone (NCBI:39878, NCBI:P:39886)

A;Note: mutant sequence of patient with osteogenesis imperfecta type IV; the authors sug

ntrol sequence

R;Maekelae, J.K.; Vuorio, T.; Vuorio, E.  
Biochim. Biophys. Acta 1049, 171-176, 1990

A;Title: Growth-dependent modulation of type I collagen production and mRNA levels in cu

A;Reference number: S10768; MUID:90304220; PMID:2364107

A;Accession: S10768

A;Molecule type: mRNA

A;Residues: 960-1021, 'L', 1023-1188, 'D', 1190-1197, 'S', 1199-1356 <MAE>

A;Cross-references: UNIPARC:UPI000016A71D; EMBL:X55255; NID:g30101; PIDN:CAA39142.1; PID

A;Experimental source: fibroblast cell culture

R;Myers, J.C.; Chu, M.L.; Faro, S.H.; Clark, W.J.; Prockop, D.J.; Ramirez, F.  
Proc. Natl. Acad. Sci. U.S.A. 78, 3516-3520, 1981

A;Title: Cloning a cDNA for the pro-alpha2 chain of human type I collagen.

A;Reference number: A18855; MUID:81273090; PMID:6267597

A;Accession: A18855

A;Molecule type: mRNA

A;Residues: 964-979, 'V', 981-1018, 'Q', 1020 <MYE>

A;Cross-references: UNIPARC:UPI0000173BA0; GB:J00114; NID:g180393; PIDN:AAA51996.1; PID:

A;Note: 1019-Leu was also found

R;Wenstrup, R.J.; Cohn, D.H.; Cohen, T.; Byers, P.H.  
J. Biol. Chem. 263, 7734-7740, 1988

A;Title: Arginine for glycine substitution in the triple-helical domain of the products

A;Reference number: I55285; MUID:88227975; PMID:2897363

A:Accession: I5285  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1090-1107 <WEN1>  
A:Cross-references: UNIPARC:UPI000016A612; GB:M22816; NID:g179602; PIDN:AAA51844.1; PID:  
A:Accession: I70059  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1090-1101,'R',1103-1107 <WEN2>  
A:Cross-references: UNIPARC:UPI000016A613; GB:M22817; NID:g179606; PIDN:AAA51846.1; PID:  
A>Note: mutant sequence from a patient with osteogenesis imperfecta type IV

Query Match 51.2%; Score 42; DB 1; Length 1366;  
Best Local Similarity 72.7%; Pred. No. 1.1e+02;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 KKVWLGETSSA 14  
| ||||| :  
Db 1224 KHVWLGETINA 1234

RESULT 20  
T02433  
DNA binding protein EREBP-3 - common tobacco  
C:Species: Nicotiana tabacum (common tobacco)  
C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 09-Jul-2004  
C:Accession: T02433  
R:Ohme-Takagi, M.; Shinshi, H.  
Plant Cell 7, 173-182, 1995  
A>Title: Ethylene-inducible DNA binding proteins that interact with an ethylene responsive  
A:Reference number: Z14671; MUID:95276459; PMID:7756828  
A:Accession: T02433  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-225 <OHM>  
A:Cross-references: UNIPROT:Q40477; UNIPARC:UPI000009FB50; EMBL:D38124; NID:g790360; PID:  
A:Experimental source: strain BY4; tissue-type leaf

Query Match 50.0%; Score 41; DB 2; Length 225;  
Best Local Similarity 50.0%; Pred. No. 29;  
Matches 11; Conservative 1; Mismatches 2; Indels 8; Gaps 2;

Qy 2 PGKK--VWLG-----ETSSAY 15  
| ||| ||| : ||  
Db 45 PGKSRVWLGTFTAEAAKAY 66

RESULT 21  
AH2866  
S-methyltransferase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 09-Jul-2004  
C:Accession: AH2866  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; L.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, J.  
ster, E.W.  
A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AH2866  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-306 <KUR>  
A:Cross-references: UNIPROT:Q8UCX1; UNIPARC:UPI00000D1E8F; GB:AE008688; PIDN:AAL43350.1;  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu2362  
A:Map position: circular chromosome

Query Match 50.0%; Score 41; DB 2; Length 306;  
Best Local Similarity 66.7%; Pred. No. 39;

Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSS 13  
| | | | | | |  
Db 135 PSVDVWLGETLS 146

RESULT 22  
E97643  
msh protein [imported] - Agrobacterium tumefaciens (strain C58, Cereon)  
C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Jul-2004  
C:Accession: E97643  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
A:Reference number: A97359; MUID:21608551; PMID:11743194  
A:Accession: E97643  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-306 <KUR>  
A:Cross-references: UNIPROT:Q8UCX1; UNIPARC:UPI000000DIE8F; GB:AE007869; PIDN:AAK98102.1;  
C:Genetics:  
A:Gene: AGR\_C\_4290  
A:Map position: circular chromosome

Query Match 50.0%; Score 41; DB 2; Length 306;  
Best Local Similarity 66.7%; Pred. No. 39;  
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSS 13  
| | | | | | |  
Db 135 PSVDVWLGETLS 146

RESULT 23  
S36444  
hygromycin phosphotransferase - Pseudomonas pseudomallei  
C:Species: Pseudomonas pseudomallei  
C:Date: 19-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 09-Jul-2004  
C:Accession: S36444  
R;Penalzoa-Vazquez, A.; Herrera-Estrella, L.; Bailey, A.M.  
submitted to the EMBL Data Library, July 1993  
A:Description: Cloning and sequencing of the genes involved in glyphosate catabolism by Pseudomonas pseudomallei  
A:Reference number: S36444  
A:Accession: S36444  
A:Molecule type: DNA  
A:Residues: 1-420 <PEN>  
A:Cross-references: UNIPROT:Q52501; UNIPARC:UPI000000BID35; EMBL:X74325; NID:9439726; PIDN:BA000019;  
A:Experimental source: strain 22

Query Match 50.0%; Score 41; DB 2; Length 420;  
Best Local Similarity 53.8%; Pred. No. 53;  
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSS 14  
| | | | | | |  
Db 383 PGPKTWAGDGSQA 395

RESULT 24  
JC7506  
heparanase protein 2a - human  
C:Species: Homo sapiens (man)  
C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 09-Jul-2004  
C:Accession: JC7506  
R;McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hircock, M.; Pao, B.; Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000  
A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family member  
A:Reference number: JC7506  
A:Accession: JC7506  
A:Molecule type: mRNA

A:Residues: 1-480 <MCK>  
A:Cross-references: UNIPROT:Q9HB39; UNIPARC:UPI000003E88A; GB:AF282885  
C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and therapeutic properties.  
C:Genetics:  
A:Gene: hpa2a  
A:Map position: 10q23-10q24  
C:Keywords: heparin binding; membrane bound

Query Match 50.0%; Score 41; DB 2; Length 480;  
Best Local Similarity 85.7%; Pred. No. 60;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PGKKVWL 8  
| | | | |  
Db 261 PGKKIWL 267

RESULT 25  
AH2248  
proteinase [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C:Accession: AH2248  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.; DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AH2248  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-615 <KUR>  
A:Cross-references: UNIPROT:Q8YRA5; UNIPARC:UPI000000CE81A; GB:BA000019; PIDN:BA075242.1;  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr3543

Query Match 50.0%; Score 41; DB 2; Length 615;  
Best Local Similarity 77.8%; Pred. No. 77;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 PGKKVWLGE 10  
| | | | |  
Db 112 PAKPWLGE 120

RESULT 26  
AD0125  
exodeoxyribonuclease V (EC 3.1.11.5) beta chain [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AD0125  
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Fildes, I.; M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, H.; Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AD0125  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1220 <KUR>  
A:Cross-references: UNIPROT:Q8ZH88; UNIPARC:UPI000000CD786; GB:AL590842; PIDN:CAC89863.1;  
C:Genetics:  
A:Gene: recB  
C:Superfamily: exodeoxyribonuclease V 135K chain

Query Match 50.0%; Score 41; DB 2; Length 1220;  
Best Local Similarity 66.7%; Pred. No. 1.5e+02;  
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 4 KKWVLGETSSAY 15
| | | | |
DB 1103 KSNWLGEDSRAY 1114

RESULT 27
AE3301
2-deoxy-D-glucanate 3-dehydrogenase (EC 1.1.1.125) [imported] - Brucella melitensis (strain
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C;Accession: AE3301
R;DelVecchio, V.G.; Kaparatil, V.; Redkar, R.J.; Patra, G.; Mujter, C.; Los, T.; Ivanova,
.; Mazur, M.; Golezman, B.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letesee
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A;Reference number: AD3252; PMID:11756688
A;Accession: AE3301
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-122 <KUR>
A;Cross-references: UNIPROT:Q8YIP8; UNIPARC:UPI0000057C76; GB:AE008917; PIDN:AAL51576.1;
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI0395
A;Map position: I
C;Keywords: oxidoreductase

Query Match 48.8%; Score 40; DB 2; Length 122;
Best Local Similarity 58.3%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 GKKVWLGETSSA 14
| | | | |
DB 66 GKKAWAGEKGEA 77

RESULT 28
S64937
probable membrane protein YLR101c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein L2705
C;Species: Saccharomyces cerevisiae
C;Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 09-Jul-2004
C;Accession: S64937
R;Messenguy, F.; Dubois, E.; Vierendeels, F.; Scherens, B.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64935
A;Accession: S64937
A;Molecule type: DNA
A;Residues: 1-131 <MES>
A;Cross-references: UNIPROT:Q08027; UNIPARC:UPI000006C33C; EMBL:Z73272; NID:gi360482; PI
A;Experimental source: strain S288C
C;Genetics:
A;Gene: MIPS:YLR101c
A;Cross-references: SGD:S0004091
A;Map position: 12R
C;Superfamily: Saccharomyces cerevisiae probable membrane protein YLR101c
C;Keywords: transmembrane protein
F;45-62/Domain: transmembrane #status predicted <TM>

Query Match 48.8%; Score 40; DB 2; Length 131;
Best Local Similarity 54.5%; Pred. No. 25;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 KKWVLGETSSAY 15
| | | | |
DB 96 KKWIGQSSYYV 106

RESULT 29
T52019
ethylene responsive element binding factor 4 [validated] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 09-Jul-2004

```

```

C;Accession: T52019
R;Fujimoto, S.Y.; Ohta, M.; Ugui, A.; Shinahi, H.; Ohme-Takagi, M.
Plant Cell 12, 393-404, 2000
A;Title: Arabidopsis ethylene responsive element binding factors act as transcriptional
A;Reference number: Z25893
C;Accession: T52019
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-222 <FUJ>
A;Cross-references: UNIPROT:O80340; UNIPARC:UPI000000140F; EMBL:AB008106; PIDN:BAA32421.1
C;Genetics:
A;Gene: ERF-4
C;Function:
A;Description: can act as transcriptional repressor in leaves [validated, MUID:20181733]

Query Match 48.8%; Score 40; DB 2; Length 222;
Best Local Similarity 50.0%; Pred. No. 41;
Matches 11; Conservative 1; Mismatches 2; Indels 8; Gaps 2;

QY 2 PGKK--VMLG-----ETSSAY 15
| | | | |
DB 43 PGKTRVVLGTFTAEAAAY 64

RESULT 30
G97553
sugar fermentation stimulation protein homolog [imported] - Agrobacterium tumefaciens (st
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C;Accession: G97553
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: G97553
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-239 <KUR>
A;Cross-references: UNIPROT:P58429; UNIPARC:UPI00001358C6; GB:AE007869; PIDN:AAK87384.1;
C;Genetics:
A;Gene: AGR_C 2956
A;Map position: circular chromosome
C;Superfamily: sugar fermentation stimulation protein

Query Match 48.8%; Score 40; DB 2; Length 239;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSS 13
| | | | |
DB 47 PGSRWLSEHDS 58

RESULT 31
AI2773
sugar fermentation stimulation protein [imported] - Agrobacterium tumefaciens (strain C58
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C;Accession: AI2773
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan
; Kap, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AI2773
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-239 <KUR>
A;Cross-references: UNIPROT:P58429; UNIPARC:UPI00001358C6; GB:AE008688; PIDN:AAL42607.1;

```

```
A;Experimental source: strain C58 (Dupont)
C;Gene: sfsa
A;Map position: circular chromosome
C;Superfamily: sugar fermentation stimulation protein

Query Match      48.8%; Score 40; DB 2; Length 239;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSS 13
   ||::|||
Db 47 PGSRIWLSEHDS 58

RESULT 32
B82979
hypothetical protein PA5340 [imported] - Pseudomonas aeruginosa (strain PA01)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C;Accession: B82979
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: B82979
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-243 <STO>
A;Cross-references: UNIPROT:Q9HTL8; UNIPARC:UPI00000C5FB0; GB:AE004946; GB:AE004091; NID:10382966
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA5340

Query Match      48.8%; Score 40; DB 2; Length 243;
Best Local Similarity 77.8%; Pred. No. 45;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PGKKVWLGE 10
   ||::|||
Db 234 PGKRVWLLE 242

RESULT 33
E72623
probable autoantigen APE1445 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 05-Oct-2004
C;Accession: E72623
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kikuchi, Y.
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix strain K1
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: E72623
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-276 <RAW>
A;Cross-references: UNIPROT:Q9YC05; UNIPARC:UPI000005DPF8; DBJ:AP0000061; NID:gs104821;
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1445
C;Superfamily: Exosome complex 3'-5' exonuclease

Query Match      48.8%; Score 40; DB 2; Length 276;
Best Local Similarity 75.0%; Pred. No. 51;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPKKVWL 8
   ||::|||
Db 132 RPGEKVV 139
```

## RESULT 34

T02434

DNA binding protein EREBP-4 - common tobacco

C;Species: Nicotiana tabacum (common tobacco)

C;Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 31-Dec-2004

C;Accession: T02434

R;Ohme-Takagi, M.; Shinshi, H.

Plant Cell 7, 173-182, 1995

A;Title: Ethylene-inducible DNA binding proteins that interact with an ethylene responsive element

A;Reference number: Z14671; MUID:95276459; PMID:7756828

A;Accession: T02434

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-291 &lt;OHM&gt;

A;Cross-references: UNIPROT:Q40478; UNIPARC:UPI000009DFAD; EMBL:D38125; NID:9790361; PID:

A;Experimental source: strain BY4; tissue-type leaf

Query Match 48.8%; Score 40; DB 2; Length 291;

Best Local Similarity 42.9%; Pred. No. 54;

Matches 9; Conservative 2; Mismatches 4; Indels 6; Gaps 1;

QY 1 RPKKVWLG-----ETSSAY 15

||::|||

Db 165 RKGTRVWLGTDTAIEAAKAY 185

## RESULT 35

T09544

phosphoprotein phosphatase (EC 3.1.3.16), catalytic beta chain - alfalfa

C;Species: Medicago sativa (alfalfa)

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004

C;Accession: T09544

R;Visi, E.; Csordas Toth, E.; Kovacs, I.; Magyar, G.; Horvath, G.; Bagossi, P.; Gergely,

Arch. Biochem. Biophys. 360, 206-214, 1998

A;Title: Protein phosphatase 1 catalytic subunit isoforms from alfalfa: Biochemical characterization

A;Reference number: Z16730; MUID:99068922; PMID:9851832

A;Accession: T09544

A;Molecule type: mRNA

A;Residues: 1-326 &lt;VIS&gt;

A;Cross-references: UNIPROT:O65844; UNIPARC:UPI00000A4B32; EMBL:AJ002485; NID:g3176071; I:

C;Genetics:

A;Gene: PPI beta

C;Function:

A;Description: catalyzes hydrolysis of peptidyl-phosphoserine or -phosphothreonine to release

C;Superfamily: serine/threonine protein phosphatase; phosphoesterase core homology; phospho

C;Keywords: phosphoric monoester hydrolase; serine/threonine-specific phosphatase

F;26-285/Domain: phosphoprotein phosphatase homology &lt;PPP&gt;

F;54-122/Domain: phosphoesterase core homology &lt;PEC&gt;

Query Match 48.8%; Score 40; DB 2; Length 326;

Best Local Similarity 72.7%; Pred. No. 50;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RPKKVWLGET 11

||::|||

Db 19 RPGEKVLSET 29

## RESULT 36

JC5191

contractile tail sheath protein - Pseudomonas aeruginosa phage PS17

C;Species: Pseudomonas aeruginosa phage PS17

C;Date: 20-Feb-1997 #sequence\_revision 27-Feb-1997 #text\_change 09-Jul-2004

C;Accession: JC5191; JC4865

R;Sasaki, T.; Shinomiya, T.; Kumazaki, T.; Mohri, N.; Ishii, S.; Arisaka, F.

Res. Commun. Biochem. Cell Mol. Biol. 1, 93-107, 1997

A;Title: Nucleotide sequences of the contractile tail sheath and tube genes of bacteriophage

A;Reference number: JC5191

A;Accession: JC5191

A;Molecule type: DNA

A;Residues: 1-386 &lt;SA2&gt;

A;Cross-references: UNIPROT:Q38068; UNIPARC:UPI00009BCD6; DDBJ:D26449; NID:g452162; PID  
C;Genetics:  
A;Gene: FI  
C;Keywords: tail protein  
F;2-386/Product: tail sheath protein #status predicted <MAT>  
  
Query Match 48.8%; Score 40; DB 2; Length 386;  
Best Local Similarity 66.7%; Pred. No. 71;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 2 PGKKVWLGETSS 13  
||:|||||  
Db 183 PGQVWNTETSS 194  
  
RESULT 37  
S73012  
polyketide synthase pksB - Mycobacterium leprae  
N;Alternate names: L518 Fl 1 protein  
C;Species: Mycobacterium leprae  
C;Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C;Accession: S73012  
R;Smith, D.R.; Robison, K.  
submitted to the EMBL Data Library, November 1993  
A;Description: Mycobacterium leprae cosmid L518.  
A;Reference number: S72591  
A;Accession: S73012  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-958 <SMI>  
A;Cross-references: UNIPROT:Q49931; UNIPARC:UPI00000BD072; EMBL:U00023; NID:g467194; PID  
C;Genetics:  
A;Start codon: GTG  
C;Keywords: carrier protein  
F;572-767/Domain: short-chain alcohol dehydrogenase homology <SAD>  
F;843-914/Domain: acyl carrier protein homology <ACP>  
  
Query Match 48.8%; Score 40; DB 2; Length 958;  
Best Local Similarity 60.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 4 KKVWLGETSS 13  
||:|||||:  
Db 286 KQVWLGDTAT 295  
  
RESULT 38  
F96501  
hypothetical protein F28H19.4 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C;Accession: F96501  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: F96501  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1014 <STO>  
A;Cross-references: UNIPROT:Q9MAR9; UNIPARC:UPI00000A7CAF; GB:AE005173; NID:g7523671; PI  
C;Genetics:  
A;Gene: F28H19.4  
A;Map position: 1  
  
Query Match 48.8%; Score 40; DB 2; Length 1014;

Best Local Similarity 41.7%; Pred. No. 1.8e+02;  
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 2 PGKKVWLGETSS 13  
||:|||||:  
Db 843 PGQIWMGKSDS 854  
  
RESULT 39  
WPMF12  
125K protein - alfalfa mosaic virus  
N;Contains: ATP-dependent helicase (EC 3.6.1.-); mRNA (guanine-N7-)-methyltransferase (EC  
C;Species: alfalfa mosaic virus, AMV  
C;Date: 14-Nov-1983 #sequence\_revision 14-Nov-1983 #text\_change 09-Jul-2004  
C;Accession: A04197  
R;Cornelissen, B.J.C.; Brederode, F.T.; Moormann, R.J.M.; Bol, J.F.  
Nucleic Acids Res. 11, 1253-1265, 1983  
A;Title: Complete nucleotide sequence of alfalfa mosaic virus RNA 1.  
A;Reference number: A04197; MUID:83143345; PMID:6298738  
A;Accession: A04197  
A;Molecule type: mRNA  
A;Residues: 1-1126 <COR>  
A;Cross-references: UNIPROT:P03589; UNIPARC:UPI0000137F7B; GB:L00163; GB:J02000; NID:g331  
C;Genetics:  
A;Map position: segment 1  
C;Superfamily: cucumber mosaic virus RNA 1 protein  
C;Keywords: hydrolase; methyltransferase; mRNA capping; nucleotide binding; P-loop; S-ad  
F;99-188/Domain: methyltransferase #status predicted <MTF>  
F;838-845/Region: nucleotide-binding motif A (P-loop)  
F;845-1096/Domain: helicase #status predicted <HHS>  
  
Query Match 48.8%; Score 40; DB 1; Length 1126;  
Best Local Similarity 42.9%; Pred. No. 2e+02;  
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
  
Qy 2 PGKKVWLGETSSAY 15  
||:|||||:  
Db 752 PKXNTWVGPTARSY 765  
  
RESULT 40  
A87204  
polyketide synthase [imported] - Mycobacterium leprae  
C;Species: Mycobacterium leprae  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
C;Accession: A87204  
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq  
A;Title: Massive gene decay in the leprosy bacillus.  
A;Reference number: A86909; MUID:21128732; PMID:11234002  
A;Accession: A87204  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1871 <STO>  
A;Cross-references: UNIPROT:Q9Z5K6; UNIPARC:UPI00000D42DC; GB:AL450380; NID:gl3093966; PJ  
C;Genetics:  
A;Gene: ML2357  
C;Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoacy  
homology; [acyl-carrier-protein] S-malonyltransferase homology  
C;Keywords: carrier protein  
  
Query Match 48.8%; Score 40; DB 2; Length 1871;  
Best Local Similarity 60.0%; Pred. No. 3.3e+02;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 4 KKVWLGETSS 13  
||:|||||:  
Db 1199 KQVWLGDTAT 1208

```

RESULT 41
AB2718
conserved hypothetical protein Atull147 [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AB2718
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AB2718
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-115 <KUR>
A:Cross-references: UNIPROT:Q8UG92; UNIPARC:UPI000016459D; GB:AE008688; PIDN:AAL42160.1;
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atull147
A:Map position: circular chromosome
C:Superfamily: Bacillus subtilis hypothetical protein yung

Query Match 47.6%; Score 39; DB 2; Length 115;
Best Local Similarity 58.3%; Pred. No. 32;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 KKVLGGETSSAY 15
||| ||||| :
DB 13 KKGWSGETSNLW 24

RESULT 42
F97499
hypothetical protein AGR_C_2123 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 31-Dec-2004
C:Accession: F97499
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: F97499
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-182 <KUR>
A:Cross-references: UNIPROT:Q8UG92; UNIPARC:UPI00000DIA64; GB:AE007869; PIDN:AAK86951.1;
C:Genetics:
A:Gene: AGR_C_2123
A:Map position: circular chromosome
C:Superfamily: Bacillus subtilis hypothetical protein yung

Query Match 47.6%; Score 39; DB 2; Length 182;
Best Local Similarity 58.3%; Pred. No. 50;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 KKVLGGETSSAY 15
||| ||||| :
DB 80 KKGWSGETSNLW 91

RESULT 43
A86461
hypothetical protein F14W2.12 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: A86461
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.

```

```

Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzbeg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A86461
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-184 <SYO>
A:Cross-references: UNIPROT:Q9LQ28; UNIPARC:UPI00000A7C94; GB:AE005172; NID:g9665094; PII
C:Genetics:
A:Map position: 1

Query Match 47.6%; Score 39; DB 2; Length 184;
Best Local Similarity 53.3%; Pred. No. 51;
Matches 8; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 2 PGKK--VWLGETSSA 14
||||| :
DB 38 PGKTRIWLSYETA 52

RESULT 44
H64968
acetyl CoA acetyltransferase - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: H64968; I69646; I69656
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Coi
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H64968
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-196 <BLAT>
A:Cross-references: UNIPROT:P37750; UNIPARC:UPI000003EB11; GB:AE000294; GB:U00096; NID:G
A:Experimental source: strain K-12, substrain MG1655
R:Yao, Z.; Valvano, M.A.
J. Bacteriol. 176, 4133-4143, 1994
A:Title: Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (rfb)
erotypes Y and 4a.
A:Reference number: I55053; MUID:94292434; PMID:7517390
A:Accession: I69646
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-167, 'LFRKYCHC', 177-186, 'IMR', 190-196 <RES>
A:Cross-references: UNIPARC:UPI000016F12B; EMBL:U03041; NID:g501028; PIDN:AAC31635.1; PII
R:Stevenson, G.; Neal, B.; Liu, D.; Hobbs, M.; Packer, N.H.; Batley, M.; Redmond, J.W.; I
J. Bacteriol. 176, 4144-4156, 1994
A:Title: Structure of the O antigen of Escherichia coli K-12 and the sequence of its rfb
A:Reference number: I55054; MUID:94292435; PMID:7517391
A:Accession: I69656
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-167, 'LFRKYCHC', 177-186, 'IMR', 190-196 <RE2>
A:Cross-references: UNIPARC:UPI000016F12B; EMBL:U09876; NID:g508236; PID:g508245
C:Genetics:
A:Gene: yefH
A:Map position: 45 min
C:Superfamily: galactoside acetyltransferase

Query Match 47.6%; Score 39; DB 2; Length 196;
Best Local Similarity 75.0%; Pred. No. 54;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 GKQVWLGE 10
||| :
DB 136 GQVWLGE 143

```

```
RESULT 45
T17997
hypothetical protein A495R - Chlorella virus PBCV-1
C:Species: Chlorella virus PBCV-1
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T17997
R:Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z18806
A:Accession: T17997
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-221 <GRA>
A:Cross-references: UNIPROT:Q98545; UNIPARC:UPI00000F8EBB; EMBL:U42580; NID:g4028896; P1
A:Experimental source: specific host Chlorella strain NC64A
C:Genetics:
A:Note: A495R
C:Superfamily: Chlorella virus PBCV-1 hypothetical protein A315L

Query Match 47.6%; Score 39; DB 2; Length 221;
Best Local Similarity 53.8%; Pred. No. 60;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 RPQKKVWLGETSS 13
:|:|:|:|:|
Db 108 KKGKNKWLKTKHS 120

RESULT 46
F64901
ABC-type transport protein b1483 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 05-Oct-2004
C:Accession: F64901
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: F64901
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-308 <BLAT>
A:Cross-references: UNIPROT:P77622; UNIPARC:UPI000013A940; GB:AE0000245; GB:U000096; NID:g
A:Experimental source: strain K-12, substrain MGL655
C:Keywords: ATP; nucleotide binding; P-loop
F:32-219/Domain: ATP-binding cassette homology <ABC>
F:49-56/Region: nucleotide-binding motif A (P-loop)

Query Match 47.6%; Score 39; DB 2; Length 308;
Best Local Similarity 54.5%; Pred. No. 84;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 PGKKVWLGETS 12
|:|:|:|:|:|
Db 16 PARKNWLKTKT 26

RESULT 47
B85728
hypothetical protein Z2227 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 05-Oct-2004
C:Accession: B85728
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouasis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85728
A:Status: preliminary
```

```
A:Molecule type: DNA
A:Residues: 1-308 <STO>
A:Cross-references: UNIPROT:Q8X4Z6; UNIPARC:UPI00000D0D46; GB:AE005174; NID:g12515198; P1
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z2227

Query Match 47.6%; Score 39; DB 2; Length 308;
Best Local Similarity 54.5%; Pred. No. 84;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 PGKKVWLGETS 12
|:|:|:|:|:|
Db 16 PARKNWLKTKT 26

RESULT 48
G90889
hypothetical protein ECs2087 [imported] - Escherichia coli (strain O157:H7, substrain RM
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 05-Oct-2004
C:Accession: G90889
R:Havashi, T.; Makino, K.; Kurokawa, K.; Ohnishi, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G90889
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-308 <HAY>
A:Cross-references: UNIPROT:Q8X4Z6; UNIPARC:UPI00000D0D46; GB:BA0000007; PIDN:BAB35510.1;
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs2087

Query Match 47.6%; Score 39; DB 2; Length 308;
Best Local Similarity 54.5%; Pred. No. 84;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 PGKKVWLGETS 12
|:|:|:|:|:|
Db 16 PARKNWLKTKT 26

RESULT 49
XNBYUG
UDPglucose-hexose-1-phosphate uridylyltransferase (EC 2.7.7.12) - yeast (Saccharomyces ce
N:Alternate names: galactose-1-phosphate uridylyltransferase; protein YBR018c; protein YBR
C:Species: Saccharomyces cerevisiae
C:Date: 17-Mar-1987 #sequence_revision 09-Sep-1994 #text_change 05-Oct-2004
C:Accession: S45873; S50813; S05811; A00720; S18725; S18757; S24918; S50322
R:Enian, K.D.; Koettler, P.; Rose, M.; Li, Z.; Thermann, R.; Brendel, M.; Baur, A.; Bolet
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45862
A:Accession: S45873
A:Molecule type: DNA
A:Residues: 1-366 <ENT>
A:Cross-references: UNIPROT:P08431; UNIPARC:UPI00001682B8; EMBL:Z35887; NID:g536219; PIDN
A:Experimental source: strain S288C
R:Schaaff-Gerstenschlaeger, I.; Schindewolf, T.; Lehnert, W.; Rose, M.; Zimmermann, F.K.
Yeast 11, 79-83, 1995
A:Title: Sequence and functional analysis of a 7.2 kb fragment of Saccharomyces cerevisiae
A:Reference number: S50812; MUID:95282516; PMID:7762304
A:Accession: S50813
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-366 <SCW>
A:Cross-references: UNIPARC:UPI00001682B8; EMBL:X81324; NID:g587572; PIDN:CAA57105.1; P1
A:Experimental source: strain S288C
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1994
R:Tajima, M.; Nogi, Y.; Fukasawa, T.
Yeast 1, 67-77, 1985
```



A;Title: Primary structure of the *Saccharomyces cerevisiae* GAL7 gene.  
 A;Reference number: S05811; MUID:89131252; PMID:2851900  
 A;Accession: S05811  
 A;Molecule type: DNA  
 A;Residues: 1-84,'S',87-266,'A',268-344,'I',346-366 <TAJ>  
 A;Cross-references: UNIPARC:UPI0000168C2E; EMBL:M12348; NID:gl71559; PIDN:AAA34627.1; PI  
 R;Citron, B.A.; Donelson, J.E.  
 J. Bacteriol. 158, 269-278, 1984  
 A;Title: Sequence of the *Saccharomyces* GAL region and its transcription in vivo.  
 A;Reference number: A91795; MUID:84185433; PMID:6715281  
 A;Accession: A00720  
 A;Molecule type: DNA  
 A;Residues: 1-10,'Y',12-57,'H',59-184 <CIT>  
 A;Cross-references: UNIPARC:UPI0000172706; EMBL:X01752; NID:gl71561; PIDN:AAA34628.1; PI  
 R;Nogi, Y.; Fukasawa, T.  
 Nucleic Acids Res. 11, 8555-8568, 1983  
 A;Title: Nucleotide sequence of the transcriptional initiation region of the yeast GAL7  
 A;Reference number: S18725; MUID:84169499; PMID:6324089  
 A;Accession: S18725  
 A;Status: translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-21 <NOG>  
 A;Cross-references: UNIPARC:UPI0000172707; EMBL:X00215  
 A;Accession: S18757  
 A;Molecule type: protein  
 A;Residues: 2-8 <NOG2>  
 A;Cross-references: UNIPARC:UPI0000172708  
 C;Genetics:  
 A;Gene: SGD:GAL7; MIPS:YBR018C  
 A;Cross-references: SGD:S0000222; MIPS:YBR018C  
 A;Map position: 2R  
 C;Function:  
 A;Description: galactose metabolism; nucleotidyltransferase  
 C;Superfamily: galactose-1-phosphate uridylyltransferase  
 C;Keywords: galactose metabolism; nucleotidyltransferase  
 F;2-366/Product: UDPglucose-hexose-1-phosphate uridylyltransferase #status experimental

Query Match 47.6%; Score 39; DB 1; Length 366;  
 Best Local Similarity 50.0%; Pred. No. 99;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 KKVLGETSSAY 15  
 |:|:|:|:  
 DB 32 KRPWLGQEAAY 43

RESULT 50  
 B70579  
 probable cell division protein FtsZ - *Mycobacterium tuberculosis* (strain H37RV)  
 C;Species: *Mycobacterium tuberculosis*  
 C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
 C;Accession: B70579  
 R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A;Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A;Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
 A;Reference number: A70500; MUID:98295987; PMID:9634230  
 A;Accession: B70579  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-379 <COL>  
 A;Cross-references: UNIPROT:O08378; UNIPARC:UPI000012AD2A; GB:Z95388; GB:AL123456; NID:9  
 A;Experimental source: strain H37RV  
 C;Genetics:  
 A;Gene: ftsZ  
 C;Superfamily: cell division protein ftsZ

Query Match 47.6%; Score 39; DB 2; Length 379;  
 Best Local Similarity 50.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSSAY 15  
 |:|:|:|:|:  
 DB 317 PGRKPVWGETGGAH 330

Search completed: June 5, 2006, 12:53:10  
 Job time : 44.6438 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:32:17 ; Search time 131.507 Seconds

(without alignment)  
105.510 Million cell updates/sec

Title: US-10-645-659A-9

Perfect score: 82

Sequence: 1 RPCKKWLGETSSAY 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : UniProt 7.2.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	100.0	543	1 HPSE_HUMAN	Q9Y251 homo sapien
2	82	100.0	558	2 Q33X5_SPAJD	Q33X5 spalax juda
3	82	100.0	574	2 Q33X6_SPAJD	Q33X6 spalax juda
4	82	100.0	574	2 Q33X7_9RODE	Q33X7 spalax carm
5	82	100.0	574	2 Q33X8_9RODE	Q33X8 spalax gola
6	82	100.0	574	2 Q33X9_9RODE	Q33X9 spalax gali
7	77	93.9	535	1 HPSE_MOUSE	Q6Ygz1 mus musculus
8	77	93.9	536	1 HPSE_RAT	Q71rpl rattus norv
9	73	89.0	523	1 HPSE_CHICK	Q90yk5 gallus gall
10	71	86.6	533	2 Q4SYF6_TETNG	Q4syf6 tetraodon n
11	68	82.9	545	1 HPSE_BOVIN	Q9myy0 bos taurus
12	64	78.0	255	2 Q4TGC8_TETNG	Q4tgc8 tetraodon n
13	50	61.0	597	2 Q4TB80_TETNG	Q4tb80 tetraodon n
14	49	59.8	586	2 Q21PR2_9DELTA	Q21pr2 anaeromyxob
15	48	58.5	197	2 Q36KN3_WARHY	Q36kn3 marinobacte
16	48	58.5	616	2 Q3QDK4_9GNMM	Q3qdk4 shewanella
17	47	57.3	334	2 Q41J33_METBU	Q41jj3 methanococc
18	47	57.3	426	2 Q54359_STRLN	Q54359 streptomyce
19	47	57.3	3143	2 Q61AL3_9POTV	Q61al3 plum pox vi
20	47	57.3	3143	2 Q6Y3X4_9POTV	Q6y3x4 plum pox vi
21	46	56.1	64	2 Q858F7_9CAUD	Q858f7 enterobacte
22	46	56.1	157	2 Q6J9N4_ARATH	Q6j9n4 arabidopsiis
23	46	56.1	203	2 Q47T88_THEFY	Q47t88 thermobifid
24	46	56.1	827	2 Q7S5G5_NEUCR	Q7s5g5 neurospora
25	46	56.1	1180	1 EX5B_ECOLI	P08394 escherichia
26	46	56.1	1180	2 Q21XG8_SHIDS	Q21xg8 shigella bo
27	46	56.1	1180	2 Q32CAL_SHIDS	Q32cal shigella dy
28	46	56.1	1180	2 Q3YY40_SHISS	Q3yy40 shigella so
29	46	56.1	1180	2 Q2MA17_ECOLI	Q2ma17 escherichia
30	46	56.1	1180	2 Q8X6M9_ECO57	Q8x6m9 escherichia
31	46	56.1	1180	2 Q83JW0_SHIFL	Q83jw0 shigella fl

32	46	56.1	1183	2	Q8FEB3_ECOL6	Q8feb3 escherichia
33	45.5	55.5	1596	2	Q8YKV0_ANASP	Q8ykv0 anabaena sp
34	45	54.9	73	2	Q3PI29_PARDE	Q3pi29 paracoccus
35	45	54.9	134	2	Q41I77_SYNPF2	Q41i77 synchococc
36	45	54.9	164	1	DRE2H_ARATH	Q9e120 arabidopsiis
37	45	54.9	225	2	Q5BK39_RAT	Q5bk39 rattus norv
38	45	54.9	225	2	Q6P3A6_MOUSE	Q6p3a6 mus musculu
39	45	54.9	225	2	Q8C708_MOUSE	Q8c708 m 0 day neo
40	45	54.9	226	2	Q8A4U3_BACTN	Q8a4u3 bacteroides
41	45	54.9	331	2	Q3ITU3_NATPD	Q3itu3 natronomona
42	45	54.9	335	2	Q6FA89_ACTAD	Q6fa89 acinetobact
43	45	54.9	473	2	Q5XK86_XENLA	Q5xk86 xenopus lae
44	45	54.9	562	2	Q57MD5_SALCH	Q57md5 salmonella
45	45	54.9	562	2	Q5PE41_SALPA	Q5pe41 salmonella
46	45	54.9	562	2	Q8ZNK5_SALTY	Q8znk5 salmonella
47	45	54.9	562	2	Q8Z592_SALTI	Q8z592 salmonella
48	45	54.9	962	2	Q5SE54_DICDI	Q5se54 dictyosteli
49	45	54.9	1869	2	Q3HF42_TRIER	Q3hf42 trichodesmi
50	44	53.7	197	2	Q36TJ3_WARHY	Q36tj3 marinobacte
51	44	53.7	246	2	Q4UIV3_9BACL	Q4ulv3 paenibacill
52	44	53.7	329	2	Q6LKM9_PROPR	Q6lkm9 photobacter
53	44	53.7	336	2	Q9SKT1_ARATH	Q9skt1 arabidopsiis
54	44	53.7	341	2	Q6MZF8_HUMAN	Q6mzf8 homo sapien
55	44	53.7	349	2	Q5ARU6_EMENI	Q5aru6 aspergillus
56	44	53.7	447	2	Q5FUI5_GLUOX	Q5fui5 gluconobact
57	44	53.7	482	1	MTTB_METTE	Q9p995 methanobarc
58	44	53.7	563	1	PTFBC_ECOLI	P20966 e pts syste
59	44	53.7	563	2	Q31YX5_SHIDS	Q31yx5 shigella bo
60	44	53.7	563	2	Q32E63_SHIDS	Q32e63 shigella dy
61	44	53.7	563	2	Q32039_SHISS	Q32039 shigella so
62	44	53.7	563	2	Q8XE91_ECO57	Q8xe91 escherichia
63	44	53.7	563	2	Q8FFT4_ECOL6	Q8fft4 escherichia
64	44	53.7	563	2	Q83QW2_SHIFL	Q83qw2 shigella fl
65	44	53.7	592	2	Q6P3X8_HUMAN	Q6p3x8 homo sapien
66	44	53.7	810	2	Q2UC74_ASPOR	Q2uc74 aspergillus
67	44	53.7	819	2	Q418Y0_GIBZE	Q418y0 gibberella
68	44	53.7	834	2	Q5SJJ7_CRYNE	Q5sjj7 cryptococcu
69	44	53.7	834	2	Q5KA04_CRYNE	Q5ka04 cryptococcu
70	44	53.7	893	2	Q9Y1Y3_9METZ	Q9y1y3 ephydatia f
71	44	53.7	1016	2	Q5FT91_GLUOX	Q5ft91 gluconobact
72	43.5	53.0	382	2	Q4K3M2_PSEF5	Q4k3m2 pseudomonas
73	43.5	53.0	1187	2	Q93284_FUGRU	Q93284 fugu rubrip
74	43	52.4	249	2	Q9KGF2_BACHD	Q9kgf2 bacillus ha
75	43	52.4	255	2	Q445E2_SOLUS	Q445e2 solibacter
76	43	52.4	304	2	Q7QTN7_GIALA	Q7qtn7 giardia lam
77	43	52.4	331	2	Q4IGM6_GIBZE	Q4igm6 gibberella
78	43	52.4	385	2	Q47QC3_THEFY	Q47qc3 thermobifid
79	43	52.4	386	2	Q5DBH7_SCHJA	Q5dbh7 schistosoma
80	43	52.4	410	2	Q7NZY8_CHRVO	Q7nzy8 chromobacte
81	43	52.4	469	2	Q8UWV9_9INFA	Q8uwv9 influenza a
82	43	52.4	469	2	Q2IDF0_9INFA	Q2idf0 influenza a
83	43	52.4	494	1	MTTB1_METAC	Q8tta9 methanosarc
84	43	52.4	494	1	MTTB_METBA	Q8tta9 methanosarc
85	43	52.4	494	1	MTTB_METBF	F0C0W7 methanosarc
86	43	52.4	494	1	Q95AF6_9ASPA	Q95af6 caesia cont
87	43	52.4	495	1	MTTB1_METWA	P58973 methanosarc
88	43	52.4	495	1	MTTB2_METAC	Q8ts73 methanosarc
89	43	52.4	495	1	MTTB2_METWA	P58974 methanosarc
90	43	52.4	766	2	Q4DKD9_TRYCR	Q4dkd9 trypanosoma
91	43	52.4	766	2	Q4CVY5_TRYCR	Q4cvy5 trypanosoma
92	43	52.4	871	2	Q37YK6_SPHAR	Q37yk6 novosphingo
93	43	52.4	877	2	Q2UGU9_ASPOR	Q2ugu9 aspergillus
94	43	52.4	887	2	Q6NRB9_XENLA	Q6nrb9 xenopus lae
95	43	52.4	1022	2	Q9LD58_ARATH	Q9ld58 arabidopsiis
96	43	52.4	1022	2	Q2NRH7_SODGL	Q2nrh7 sodalis glo
97	43	52.4	1180	2	Q2NRH7_SODGL	Q2nrh7 sodalis glo
98	43	52.4	1181	2	Q57KC4_SALCH	Q57kc4 salmonella
99	43	52.4	1181	2	Q5PEM8_SALPA	Q5pem8 salmonella
100	43	52.4	1181	2	Q8ZMB6_SALTY	Q8zmb6 salmonella
101	43	52.4	1181	2	Q8Z419_SALTI	Q8z419 salmonella
102	43	52.4	1536	2	Q8GAL7_DICDI	Q8gal7 dictyosteli
103	43	52.4	2767	2	Q4B4A1_9BURK	Q4b4a1 polaromonas
104	42.5	51.8	580	2	Q3M2M4_ANAVT	Q3m2m4 anabaena va

105	42	51.2	73	2	Q44NM0_CHLLI	Q44nm0 chlorobium	178	41	50.0	365	2	Q6K7E6_ORYSA	Q6k7e6 oryza sativ
106	42	51.2	113	2	Q6I823_OSCAG	Q6i823 oscillatori	179	41	50.0	365	2	Q9SE28_ORYSA	Q9se28 oryza sativ
107	42	51.2	113	2	Q6I825_OSCAG	Q6i825 oscillatori	180	41	50.0	384	2	Q7XA04_SOYBN	Q7xa04 glyeine max
108	42	51.2	113	2	Q6I827_OSCAG	Q6i827 oscillatori	181	41	50.0	385	2	Q4B682_9BURK	Q4b682 polaromonas
109	42	51.2	113	2	Q6I831_OSCAG	Q6i831 oscillatori	182	41	50.0	396	2	Q7QNG0_ANOGA	Q7qng0 anopheles g
110	42	51.2	113	2	Q6I833_OSCAG	Q6i833 planktothri	183	41	50.0	407	2	Q3G9A3_9FIRM	Q3g9a3 syntrophomo
111	42	51.2	113	2	Q6I835_OSCAG	Q6i835 oscillatori	184	41	50.0	420	2	Q2S501_BURPS	Q2s501 burkholderi
112	42	51.2	118	2	Q69KQ3_ORYSA	Q69kq3 oryza sativ	185	41	50.0	455	2	Q5B577_EMENI	Q5b577 aspergillus
113	42	51.2	130	2	Q52127_92ZZZ	Q52127 plasmid pl.	186	41	50.0	469	2	Q6KV30_9INFA	Q6kv30 influenza a
114	42	51.2	134	2	Q86445_PLARU	Q86445 planktothri	187	41	50.0	469	2	Q710U6_IACKS	Q710u6 influenza a
115	42	51.2	134	2	Q88099_9CYAN	Q88099 planktothri	188	41	50.0	474	2	Q6R7A2_9HERP	Q6r7a2 ostreid her
116	42	51.2	134	2	Q798E8_PLARU	Q798e8 planktothri	189	41	50.0	490	2	Q7EYM6_ORYSA	Q7eym6 oryza sativ
117	42	51.2	134	2	Q798F3_9CYAN	Q798f3 planktothri	190	41	50.0	538	2	Q5Z367_NOCFA	Q5z367 nocardia fa
118	42	51.2	134	2	Q798F7_OSCAG	Q798f7 oscillatori	191	41	50.0	559	2	Q89F99_BRAJA	Q89f99 bradyrhizob
119	42	51.2	134	2	Q7B2U5_PLARU	Q7b2u5 planktothri	192	41	50.0	592	1	HPSE2_HUMAN	Q8hwq2 homo sapien
120	42	51.2	231	2	Q3FRF5_9BURK	Q3frf5 rhodofexax	193	41	50.0	606	2	Q2M1H9_HUMAN	Q2m1h9 homo sapien
121	42	51.2	231	2	Q871B0_NEUCR	Q871b0 neurospora	194	41	50.0	606	2	Q2NQ27_SODGL	Q2nq27 sodalis glo
122	42	51.2	233	2	Q7R214_NEUCR	Q7r214 neurospora	195	41	50.0	615	2	Q4LC08_SODGL	Q4lc08 sodalis glo
123	42	51.2	266	2	Q4E222_CHRLS	Q4e222 chromohalob	196	41	50.0	615	2	Q3M7A7_ANAVT	Q3m7a7 anabaena va
124	42	51.2	288	2	Q7XY15_CHLS6	Q7xy15 chlorarachn	197	41	50.0	615	2	Q8YRA5_ANASP	Q8yra5 anabaena sp
125	42	51.2	318	2	Q60ZH3_CAEBR	Q60zh3 caenorhabdi	198	41	50.0	664	2	Q2WSS4_CLOBE	Q2wss4 clostridium
126	42	51.2	341	1	DRE2C_ARATH	Q8lfr2 arabidopsis	199	41	50.0	699	2	Q5W7P2_APIME	Q5w7p2 apis mellif
127	42	51.2	343	2	Q6U5L6_KLEPN	Q6u5l6 klebsiella	200	41	50.0	713	2	Q7YV49_CRYPV	Q7yv49 cryptospori
128	42	51.2	366	2	Q9HUS4_PSEAE	Q9hus4 pseudomonas	201	41	50.0	737	2	Q981S7_RHILO	Q981s7 rhizobium l
129	42	51.2	371	1	Y4OS_RH1SN	P55604 rhizobium s	202	41	50.0	831	2	Q7VNB0_HAEDU	Q7vnb0 haemophilus
130	42	51.2	429	2	Q61PB3_CAEBR	Q61pb3 caenorhabdi	203	41	50.0	853	2	Q86ZNO_PODAN	Q86zn0 podospora a
131	42	51.2	454	2	Q33SJO_9GAMW	Q33sj0 shewanella	204	41	50.0	896	2	Q4WPA2_ASPFU	Q4wpa2 aspergillus
132	42	51.2	463	2	Q37TL5_SPHAR	Q37tl5 novosphingo	205	41	50.0	1162	2	Q8D2T8_WIGBR	Q8d2t8 wigglewort
133	42	51.2	493	2	Q7PH57_NOGA	Q7ph57 anopheles g	206	41	50.0	1220	2	Q667G7_YERPE	Q667g7 yersinia ps
134	42	51.2	498	2	Q3SLB7_THIDA	Q3slb7 thiobacilli	207	41	50.0	1220	2	Q8ZH88_YERPE	Q8zh88 yersinia pe
135	42	51.2	515	2	Q8T108_BOMMO	Q8t108 bombyx mori	208	41	50.0	1241	2	Q8CZY1_YERPE	Q8czy1 yersinia pe
136	42	51.2	527	2	Q9LRC8_SCUBA	Q9lrc8 scutellaria	209	41	50.0	1338	2	Q4SW64_PHACH	Q4sw64 phanerochae
137	42	51.2	530	2	Q4HY92_GIBZE	Q4hy92 gibberella	210	40.5	49.4	382	2	Q3K4G6_PSEPF	Q3k4g6 pseudomongo
138	42	51.2	591	2	Q5NNY0_CRYNE	Q5nny0 cryptococcu	211	40.5	49.4	640	2	Q3K7V1_SPHAR	Q3k7v1 novosphingo
139	42	51.2	591	2	Q5KEH9_CRYNE	Q5keh9 cryptococcu	212	40	48.8	86	2	Q2P9E5_XANOR	Q2p9e5 xanthomonas
140	42	51.2	661	2	Q7PQJ2_NOGA	Q7pqj2 anopheles g	213	40	48.8	114	2	Q7Q2Z1_LEPIC	Q7q2z1 leptospira
141	42	51.2	775	2	Q66593_AQUAE	Q66593 aquifex aeo	214	40	48.8	114	2	Q8F4U7_LEPIC	Q8f4u7 leptospira
142	42	51.2	809	2	Q53F43_HUMAN	Q53f43 homo sapien	215	40	48.8	122	2	Q8YIP8_BRUME	Q8yip8 brucella me
143	42	51.2	809	2	Q96J53_HUMAN	Q96j53 homo sapien	216	40	48.8	131	2	Q80027_YEAST	Q80027 saccharomyc
144	42	51.2	833	1	M4K1_HUMAN	Q92918 homo sapien	217	40	48.8	151	2	Q8KKU5_RHIET	Q8kku5 rhizobium e
145	42	51.2	890	2	Q8EWR8_MYCPE	Q8ewr8 mycoplasma	218	40	48.8	167	2	Q304Y6_CAPAN	Q304y6 capsicum an
146	42	51.2	921	2	Q6CF18_YARLI	Q6cf18 yarrowia li	219	40	48.8	173	2	Q891V6_BRAJA	Q891v6 bradyrhizob
147	42	51.2	1041	2	Q4QFB8_LEIMA	Q4qfb8 leishmania	220	40	48.8	174	2	Q3RZX1_CALME	Q3rxz1 raietionella m
148	42	51.2	1050	2	Q2IP55_9DELTA	Q2ip55 anaeromyxob	221	40	48.8	178	2	Q2Z181_RALSA	Q2z181 caldicellul
149	42	51.2	1083	2	Q5U045_MIMIV	Q5uq45 mimivirus	222	40	48.8	196	2	Q3IH6_PSEHT	Q3ih6 pseudosalter
150	42	51.2	1366	1	CO1A2_HUMAN	P08123 homo sapien	223	40	48.8	199	2	Q4T222_TETNG	Q4t222 tetradon n
151	42	51.2	1366	2	Q15177_HUMAN	Q15177 homo sapien	224	40	48.8	200	1	ERF9_ARATH	Q9f67 arabidopsis
152	42	51.2	1366	2	Q7Z5S6_HUMAN	Q7z5s6 homo sapien	225	40	48.8	209	2	Q8ELK0_OCEIH	Q8elk0 oceanobacil
153	42	51.2	1481	2	Q5VWL1_HUMAN	Q5vwl1 homo sapien	226	40	48.8	210	2	Q84LQ6_LYCES	Q84lq6 lycopersico
154	42	51.2	1617	2	Q4WXY0_ASPFU	Q4wxy0 aspergillus	227	40	48.8	216	2	Q2N5X0_9SPHN	Q2n5x0 erythrobaet
155	41.5	50.6	95	2	Q9IA90_ORYLA	Q9ia90 oryzias lat	228	40	48.8	218	2	Q212S8_MEDTR	Q212s8 medicago tr
156	41.5	50.6	220	2	Q2NXT6_XANOR	Q2nxt6 xanthomonas	229	40	48.8	218	2	Q3AIK7_SYNSC	Q3aik7 synchococc
157	41.5	50.6	290	2	Q5GUG6_XANOR	Q5guh6 xanthomonas	230	40	48.8	218	2	Q7V7D5_PROMM	Q7v7d5 prochloroco
158	41.5	50.6	774	2	Q4SGZ3_TETNG	Q4sgz3 tetradon n	231	40	48.8	220	2	Q6TKQ3_VITAE	Q6tkq3 vitis aesti
159	41	50.0	177	2	Q5CE78_CRYHO	Q5ce78 cryptospori	232	40	48.8	222	1	ERF4_ARATH	Q80340 arabidopsis
160	41	50.0	204	2	Q6RUR2_CAPAN	Q6rur2 capsicum an	233	40	48.8	222	2	Q310Q8_GOSHI	Q310q8 gossypium h
161	41	50.0	211	2	Q32W75_CAPAN	Q32w75 capsicum an	234	40	48.8	222	2	Q53XI2_ARATH	Q53xi2 arabidopsis
162	41	50.0	225	1	ERF4_TOBAC	Q40477 nicotiana t	235	40	48.8	222	2	Q84XB1_LYCES	Q84xb1 lycopersico
163	41	50.0	227	1	ERF4_NICSY	Q91w49 nicotiana s	236	40	48.8	226	2	Q41W90_DESHA	Q41w90 desulfitoba
164	41	50.0	249	2	Q6TKQ4_VITAE	Q6tkq4 vitis aesti	237	40	48.8	228	2	Q5UX44_HALMA	Q5ux44 haloarcula
165	41	50.0	252	2	Q2RYW8_SPHI	Q2ryw8 salinibacte	238	40	48.8	229	2	Q9WH76_9RHAB	Q9wh76 chandipura
166	41	50.0	259	2	Q5IWL7_TOBAC	Q5iwl7 nicotiana t	239	40	48.8	239	1	SFSA_AGR75	P58429 agrobacteri
167	41	50.0	273	2	Q5TIL3_FUGRU	Q5til3 fugu rubrip	240	40	48.8	243	2	Q9HTL8_PSEAE	Q9htl8 pseudomonas
168	41	50.0	280	2	Q50J61_FUGRU	Q50j61 fugu rubrip	241	40	48.8	248	2	Q3JB84_NITOC	Q3jb84 nitrosococc
169	41	50.0	306	2	Q8UCX1_AGR75	Q8ucx1 agrobacteri	242	40	48.8	250	2	Q88MN3_PSEPK	Q88mn3 pseudomonas
170	41	50.0	308	2	Q41BR1_GIBZE	Q41br1 gibberella	243	40	48.8	252	2	Q2NB03_9SPHN	Q2nb03 erythrobaet
171	41	50.0	312	2	Q94HF2_ORYSA	Q94hf2 oryza sativ	244	40	48.8	257	2	Q2RXD3_RHORU	Q2rxd3 rhodospiril
172	41	50.0	326	2	Q7XEH8_ORYSA	Q7xbh8 oryza sativ	245	40	48.8	257	2	Q57BP7_BRUAB	Q57bp7 brucella ab
173	41	50.0	326	2	Q8VXC3_ORYSA	Q8vxc3 oryza sativ	246	40	48.8	257	2	Q8FZ65_BRUSU	Q8fz65 brucella su
174	41	50.0	332	2	Q84LJ7_WHEAT	Q84lj7 triticum ae	247	40	48.8	257	2	Q4YQB9_BRUA2	Q4yqb9 brucella ab
175	41	50.0	349	2	Q8HOK1_WHEAT	Q8hok1 triticum ae	248	40	48.8	258	2	Q4PKD2_9ROSI	Q4pkd2 jatrophia cu
176	41	50.0	355	2	Q2TN80_WHEAT	Q2tn80 triticum ae	249	40	48.8	276	1	ECK2_AERPE	Q9yc05 aeropyrum p
177	41	50.0	355	2	Q7XY26_WHEAT	Q7xy26 triticum ae	250	40	48.8	282	1	ERF5_NICSY	Q91w48 nicotiana s

251	40	48.8	291	1	ERP5_TOBAC	Q40478 nicotiana t	324	39.5	48.2	383	2	Q8NLE1_CORGL	Q8nle1 corynebacte
252	40	48.8	293	2	Q52242_92ZZZ	Q52242 plasmid pvt	325	39.5	48.2	387	2	Q8FLN3_CORREF	Q8fln3 corynebacte
253	40	48.8	293	2	Q7W282_BORPA	Q7W282 bordetella	326	39.5	48.2	568	2	Q5B7S5_EMENI	Q5b7s5 aspergillus
254	40	48.8	293	2	Q7WR50_BORBR	Q7WR50 bordetella	327	39.5	48.2	810	2	Q82MJ1_STRAW	Q82mj1 streptomyce
255	40	48.8	313	2	Q4SHB8_TETNG	Q4SHB8 tetradon n	328	39.5	48.2	2257	2	Q6K777_ORYSA	Q6k777 oryza sativ
256	40	48.8	318	2	Q623H9_ORYSA	Q623h9 oryza sativ	329	39	47.6	46	2	Q9H3V6_HUMAN	Q9h3v6 homo sapien
257	40	48.8	319	2	Q6K4M2_SYRGO	Q6K4m2 streptomyce	330	39	47.6	78	2	Q30WK2_DESDG	Q30wk2 desulfovibr
258	40	48.8	323	2	Q7NDQ9_GLOVI	Q7NDq9 gloebacter	331	39	47.6	87	2	Q32HM4_SHIDS	Q32hm4 shigella dy
259	40	48.8	324	2	Q8LML9_ORYSA	Q8lml9 oryza sativ	332	39	47.6	93	2	Q2NXL1_XANOR	Q2nxl1 xanthomonas
260	40	48.8	326	2	Q65844_MEDSA	Q65844 medicago sa	333	39	47.6	93	2	Q5GUA1_XANOR	Q5gua1 xanthomonas
261	40	48.8	331	2	Q2PC80_STRAH	Q2PC80 streptomyce	334	39	47.6	95	2	Q2LQ46_9DEL1	Q2lq46 syntrophus
262	40	48.8	333	2	Q41PL9_METBU	Q41PL9 methanococc	335	39	47.6	95	2	Q6NJ54_CORDI	Q6nj54 corynebacte
263	40	48.8	341	2	Q6RZW7_VITAE	Q6rzw7 vitis aesti	336	39	47.6	96	2	Q6D9H8_ERWCT	Q6d9h8 erwina car
264	40	48.8	366	2	Q4H7M0_9DEIO	Q4H7m0 deinococcus	337	39	47.6	110	2	Q8IQ66_DROME	Q8iq66 drosophila
265	40	48.8	369	2	Q4LH6_CAPAN	Q4Lh6 capsicum an	338	39	47.6	115	2	Q8UG92_AGRF5	Q8ug92 agrobacteri
266	40	48.8	369	2	Q99CE1_BRAJA	Q99CE1 bradyrhizob	339	39	47.6	128	2	Q6MMK9_ORYSA	Q6mmk9 oryza sativ
267	40	48.8	370	2	Q4IXG6_AZOVI	Q4IXg6 azotobacter	340	39	47.6	158	2	Q8IGF2_DROPHI	Q8igf2 drosophila
268	40	48.8	372	2	Q8LGR9_LYCES	Q8lgr9 lycopersico	341	39	47.6	158	2	Q3F5W4_9BURK	Q3f5w4 burkholderi
269	40	48.8	375	2	Q829I9_STRAW	Q829i9 streptomyce	342	39	47.6	166	2	Q7Y1E5_ORYSA	Q7y1e5 oryza sativ
270	40	48.8	376	2	Q4T7U2_TETNG	Q4T7u2 tetradon n	343	39	47.6	175	2	Q7PWN1_ANOGA	Q7pwn1 anopheles g
271	40	48.8	380	2	Q4H4G2_BACCI	Q4H4g2 bacillus ci	344	39	47.6	182	2	Q7CZV8_AGRF5	Q7czv8 agrobacteri
272	40	48.8	383	2	Q2U2B8_ASPOR	Q2U2b8 aspergillus	345	39	47.6	184	2	Q9LQ28_ARATH	Q9lq28 arabidopsis
273	40	48.8	386	2	Q38068_BACTERI	Q38068 bacterioph	346	39	47.6	192	2	Q2WB59_MAGSA	Q2wb59 magnetospi
274	40	48.8	390	2	Q8GCB3_BACCI	Q8gcb3 bacillus ci	347	39	47.6	192	2	Q9AJL8_HYDTH	Q9ajl8 hydrogenoba
275	40	48.8	391	2	Q8DJ64_SYNEL	Q8dj64 synectococc	348	39	47.6	192	2	Q2ISL7_RHOPA	Q2isl7 rhodopseuo
276	40	48.8	391	2	Q9RJP0_STRCO	Q9Rjp0 streptomyce	349	39	47.6	195	2	Q8S2S7_THEHA	Q8s2s7 thellungiel
277	40	48.8	398	2	Q5FTR6_GLUOX	Q5ftr6 gluconobact	350	39	47.6	196	1	WBBJ_ECOLI	W3750 escherichia
278	40	48.8	410	2	Q3UC89_MOUSE	Q3uc89 m bone marr	351	39	47.6	197	2	Q6ZAH8_ORYSA	Q6zah8 oryza sativ
279	40	48.8	426	2	Q2P244_XANOR	Q2p244 xanthomonas	352	39	47.6	198	2	Q6L4M2_ORYSA	Q6l4m2 oryza sativ
280	40	48.8	426	2	Q3BS70_XANC5	Q3bs70 xanthomonas	353	39	47.6	203	2	Q3FQ51_9BURK	Q3fq51 rhodofera
281	40	48.8	426	2	Q4UVU6_XANC8	Q4uvu6 xanthomonas	354	39	47.6	207	2	Q4R4W9_MACFA	Q4r4w9 macaca fasc
282	40	48.8	426	2	Q8P891_XANCP	Q8p891 xanthomonas	355	39	47.6	212	2	Q7VT35_BORPE	Q7vt35 bordetella
283	40	48.8	431	2	Q3TDG5_MOUSE	Q3tdg5 mus musculu	356	39	47.6	212	2	Q7WFI2_BORBR	Q7wfi2 bordetella
284	40	48.8	442	2	Q8PJP6_XANAC	Q8pjp6 xanthomonas	357	39	47.6	221	2	Q9MS45_PBCV1	Q9ms45 paraneecium
285	40	48.8	449	2	Q2LFX4_9INFA	Q2lfx4 influenza a	358	39	47.6	231	2	Q4W6U0_NICBE	Q4w6u0 nicotiana b
286	40	48.8	449	2	Q2LFX5_9INFA	Q2lfx5 influenza a	359	39	47.6	237	2	Q9FR02_TOBAC	Q9fr02 nicotiana t
287	40	48.8	455	2	Q747L5_GEOSL	Q747l5 geobacter s	360	39	47.6	238	2	Q9ZR85_STYHA	Q9zr85 stylosanthe
288	40	48.8	493	2	Q8SYJ5_DROME	Q8syj5 drosophila	361	39	47.6	239	2	Q9ZR83_STYHA	Q9zr83 stylosanthe
289	40	48.8	493	2	Q9VMN1_DROME	Q9vmn1 drosophila	362	39	47.6	240	2	Q8H6S9_LYCES	Q8h6s9 lycopersico
290	40	48.8	496	2	Q2S2J8_9GAMW	Q2s2j8 habella che	363	39	47.6	245	2	Q8GZE9_LYCES	Q8gze9 lycopersico
291	40	48.8	511	2	Q5GZ29_XANOR	Q5gz29 xanthomonas	364	39	47.6	246	2	Q8CCT9_MOUSE	Q8cct9 mus musculu
292	40	48.8	543	2	Q7MDK8_VIBVY	Q7mdk8 vibrio vuln	365	39	47.6	247	2	Q6PHC5_BRARE	Q6phc5 brachydanio
293	40	48.8	543	2	Q8D6Q1_VIBVU	Q8d6q1 vibrio vuln	366	39	47.6	255	2	Q852K1_ENTFA	Q852k1 enterococcu
294	40	48.8	599	2	Q5HMK3_STAEQ	Q5hmk3 staphylococ	367	39	47.6	259	2	Q47L96_TEFPA	Q47l96 thermobifid
295	40	48.8	629	2	Q3VGF9_9SPHN	Q3vgf9 sphingopyxi	368	39	47.6	273	2	Q7X649_ORYSA	Q7x649 oryza sativ
296	40	48.8	668	2	Q6PG56_MOUSE	Q6pg56 mus musculu	369	39	47.6	273	2	Q9AQU3_ORYSA	Q9aqu3 oryza sativ
297	40	48.8	736	2	Q4QF12_LEIMA	Q4qf12 leishmania	370	39	47.6	277	2	Q5LQ59_SILPO	Q5lq59 silicibacte
298	40	48.8	773	2	Q93FB8_9RHOO	Q93fb8 azoarcus ev	371	39	47.6	278	2	Q6RZM8_VITAE	Q6rzm8 vitis aesti
299	40	48.8	773	2	Q93FC6_9RHOO	Q93fc6 azoarcus ev	372	39	47.6	279	2	Q4K6B4_PSEPF	Q4k6b4 pseudomonas
300	40	48.8	822	2	Q83HG9_TROW8	Q83hg9 tropheryma	373	39	47.6	279	2	Q4K5I0_PSEPK	Q4k5i0 pseudomonas
301	40	48.8	822	2	Q83GT4_TROWT	Q83gt4 tropheryma	374	39	47.6	279	2	Q48DN5_PSEPK	Q48dn5 pseudomonas
302	40	48.8	879	2	Q7TPK0_RAT	Q7tpk0 rattus norv	375	39	47.6	280	2	Q48DN5_PSE14	Q48dn5 pseudomonas
303	40	48.8	893	1	RUSC1_MOUSE	Q8b926 mus musculu	376	39	47.6	280	2	Q4ZN94_PSEU2	Q4zn94 pseudomonas
304	40	48.8	958	2	Q49931_MYCLE	Q49931 mycobacteri	377	39	47.6	280	2	Q6AC01_LEIXX	Q6ac01 leifsonia x
305	40	48.8	979	2	Q73TG7_MYCPA	Q73tg7 mycobacteri	378	39	47.6	280	2	Q87VX8_PSESM	Q87vx8 pseudomonas
306	40	48.8	988	2	Q6CPA2_KULFA	Q6cpa2 kluyveromyc	379	39	47.6	283	2	Q2UI80_ASPOR	Q2ui80 aspergillus
307	40	48.8	1014	2	Q9MAR9_ARATH	Q9mar9 arabidopsis	380	39	47.6	290	2	Q3QN11_9RHOB	Q3qn11 silicibacte
308	40	48.8	1031	2	Q2JB15_9ACTO	Q2jb15 frankia sp.	381	39	47.6	291	2	Q3W3F0_9ACTO	Q3w3f0 frankia sp.
309	40	48.8	1126	1	VIA_AMVLE	P03589 alfalfa mos	382	39	47.6	292	2	Q441N3_SOLUS	Q441n3 solibacter
310	40	48.8	1191	2	Q9RHV0_STRRO	Q9rhv0 streptomyce	383	39	47.6	292	2	Q47LKO_RALBP	Q47lko ralsconia e
311	40	48.8	1440	2	Q6VY42_9CAUD	Q6vy42 bacterioph	384	39	47.6	293	2	Q2J3E3_RHOPA	Q2j3e3 rhodopseuo
312	40	48.8	1445	2	Q93251_RANCA	Q93251 rana cateb	385	39	47.6	296	2	Q5TW01_ANOGA	Q5tw01 anopheles g
313	40	48.8	1450	2	Q9Y1B4_CYNPY	Q9y1b4 cynops pyrr	386	39	47.6	298	2	Q3HBV9_TRIER	Q3hbv9 trichodeami
314	40	48.8	1666	2	Q8LP68_CHLRE	Q8lp68 chlamydomon	387	39	47.6	308	1	YDQO_ECOLI	Y7622 escherichia
315	40	48.8	1698	2	Q94438_CHIRONO	Q94438 chironomus	388	39	47.6	308	2	Q5TW02_ANOGA	Q5tw02 anopheles g
316	40	48.8	1871	2	Q9Z5K6_MYCLE	Q9z5k6 mycobacteri	389	39	47.6	308	2	Q320S2_SHIBS	Q320s2 shigella bo
317	40	48.8	2152	2	Q9ALM5_9PSEU	Q9alm5 saccharopol	390	39	47.6	308	2	Q32G18_SHIDS	Q32g18 shigella dy
318	40	48.8	2295	2	Q83X71_STRRO	Q83x71 streptomyce	391	39	47.6	308	2	Q8X4Z6_ECO57	Q8x4z6 escherichia
319	40	48.8	3651	2	Q83X69_STRRO	Q83x69 streptomyce	392	39	47.6	310	2	Q826M4_STRAW	Q826m4 streptomyce
320	40	48.8	4928	2	Q9ALM3_9PSEU	Q9alm3 saccharopol	393	39	47.6	312	2	Q4C4C0_CROWT	Q4c4c0 crocospaer
321	40	48.8	5588	2	Q9ALM2_9PSEU	Q9alm2 saccharopol	394	39	47.6	313	2	Q2XIM4_PSEPU	Q2xim4 pseudospaer
322	39.5	48.2	124	2	Q4J8I6_SULAC	Q4j8i6 sulfolobus	395	39	47.6	316	2	Q37S31_SPHAR	Q37s31 novosphingo
323	39.5	48.2	131	1	HIS3_PYRAE	Q8zy39 pyrobaculum	396	39	47.6	318	2	Q7XSD2_ORYSA	Q7xsd2 oryza sativ

397	39	47.6	319	2	Q6ABJ4_PROAC	Q6ABJ4_propionibac	470	39	47.6	572	2	Q3B2X5_PELLDD	Q3B2x5 pelodictyon
398	39	47.6	322	2	Q4J9E0_SULAC	Q4J9e0 sulfobolus	471	39	47.6	581	2	Q4I1B5_GIBZE	Q4I1b5 gibberella
399	39	47.6	322	2	Q4KB39_PSEF5	Q4kb39 pseudomonas	472	39	47.6	594	2	Q4HVM2_GIBZE	Q4Hvm2 gibberella
400	39	47.6	326	2	Q96D47_HUMAN	Q96d47 homo sapien	473	39	47.6	609	2	Q3ZC37_BOVIN	Q3Zc37 bos taurus
401	39	47.6	327	2	Q7PSR8_ANOGA	Q7psr8 anopheles g	474	39	47.6	617	2	Q3RGC6_ACTSC	Q3Rgc6 actinobacil
402	39	47.6	327	2	Q6TXK7_LYCES	Q6txk7 lycopersico	475	39	47.6	621	2	Q41CK5_9BACI	Q41ck5 exigubacte
403	39	47.6	327	2	Q2XDC6_PSEPU	Q2xdc6 pseudomonas	476	39	47.6	622	2	Q6CGF8_YARLI	Q6Cgf8 yarrowia l1
404	39	47.6	328	2	Q7XJ22_ORISA	Q7xj22 oryza sativ	477	39	47.6	669	2	Q982K1_RHILO	Q982k1 rhitobium l1
405	39	47.6	328	2	Q7XAD5_ORISA	Q7xad5 oryza sativ	478	39	47.6	674	2	Q6BSC1_DEBHA	Q6Bsc1 debaryomyce
406	39	47.6	330	2	Q9F7C9_9RHIZ	Q9f7c9 agrobacteri	479	39	47.6	687	2	Q3ZDP9_ARAGE	Q3Zdp9 arabis gemm
407	39	47.6	343	2	Q8CFE0_MOUSE	Q8cef0 mus musculu	480	39	47.6	693	2	Q3SM98_THIDA	Q3sm98 thiobacillu
408	39	47.6	365	1	CATD_SHEEP	Q8mz88 ovis aries	481	39	47.6	703	2	Q8RXN0_ARATH	Q8rxn0 arabidopsis
409	39	47.6	365	1	GAL7_YEAST	P08431 saccharomyc	482	39	47.6	713	2	Q3PB31_PARDE	Q3pb31 paracoccus
410	39	47.6	365	2	Q3B129_PELLDD	Q3b129 pelodictyon	483	39	47.6	744	1	NSF_CRIGR	P18708 cricetulus
411	39	47.6	366	2	Q3XPQ7_9PROT	Q3xpq7 magnetococc	484	39	47.6	744	1	NSF_HUMAN	P46459 homo sapien
412	39	47.6	378	2	Q4RLR4_TETNG	Q4rlr4 tetraodon n	485	39	47.6	744	1	NSF_MOUSE	P46460 mus musculu
413	39	47.6	379	1	FTS2_MYCBO	P64171 mycobacteri	486	39	47.6	744	2	Q8N6D7_HUMAN	Q8n6d7 homo sapien
414	39	47.6	379	1	FTS2_MYCBO	P64170 mycobacteri	487	39	47.6	744	2	Q8N6D7_HUMAN	Q8n6d7 homo sapien
415	39	47.6	382	1	SUBT_BACAM	P00782 bacillus am	488	39	47.6	744	2	Q8C3R2_MOUSE	Q8c3r2 mus musculu
416	39	47.6	386	2	Q5NLV2_ZYMMO	Q5nlv2 zymomonas m	489	39	47.6	744	2	Q8C3R2_MOUSE	Q8c3r2 mus musculu
417	39	47.6	388	2	Q5AYL0_EMENI	Q5ayl0 aspergillus	490	39	47.6	744	2	Q923C6_MOUSE	Q923c6 m n-ethylma
418	39	47.6	388	2	Q3GVA9_BACTO	Q3gva9 nocardioid	491	39	47.6	744	2	Q9QUL6_RAT	Q9qul6 rattus norv
419	39	47.6	389	2	Q44H95_CHRSL	Q44h95 chromohalob	492	39	47.6	787	2	Q2VAD3_HPBUD	Q2vad3 duck hepati
420	39	47.6	390	2	Q8RD18_THETN	Q8rdl8 thermoanaer	493	39	47.6	804	2	Q3ITQ0_NATPD	Q3itq0 natronomona
421	39	47.6	406	2	Q3FN14_9BURK	Q3fn14 rhodospheax	494	39	47.6	812	2	Q4WG06_ASPFU	Q4wg06 aspergillus
422	39	47.6	407	2	Q46F29_METBA	Q46f29 methanosarc	495	39	47.6	821	2	Q2ULM7_ASPOR	Q2ulm7 aspergillus
423	39	47.6	410	2	Q9CCJ5_MYCLE	Q9ccj5 mycobacteri	496	39	47.6	828	2	Q6H503_ARATH	Q6h503 arabidopsis
424	39	47.6	420	2	Q7U360_BORPE	Q7u360 bordetella	497	39	47.6	865	2	Q8GH14_AERPU	Q8gh14 aeromonas p
425	39	47.6	420	2	Q7U371_BORPA	Q7u371 bordetella	498	39	47.6	866	2	Q9L5D5_AERHY	Q9l5d5 aeromonas h
426	39	47.6	420	2	Q7U382_BORBP	Q7u382 bordetella	499	39	47.6	867	2	Q36LU2_MARHY	Q36lu2 clostridum
427	39	47.6	424	2	Q65CX5_BACLD	Q65cx5 bacillus li	500	39	47.6	869	2	Q2WLJ0_CLOBE	Q2wlj0 clostridium
428	39	47.6	425	2	Q62NF0_BACLD	Q62nf0 bacillus li	501	39	47.6	888	2	Q4WMU3_ASPFU	Q4wmu3 aspergillus
429	39	47.6	426	2	Q8KCT1_CHLTE	Q8kct1 chlorobium	502	39	47.6	891	2	Q54RN3_DICTDI	Q54rn3 dictyosteli
430	39	47.6	427	2	Q58CY1_BOVIN	Q58cy1 bos taurus	503	39	47.6	923	2	Q4SBM6_TETNG	Q4sbm6 tetraodon n
431	39	47.6	429	1	YVFO_BACSU	Q07013 bacillus su	504	39	47.6	939	2	Q2RYX4_9SPHI	Q2ryx4 salinibacte
432	39	47.6	442	2	Q6N7G8_RHOPA	Q6n7g8 rhodopseudo	505	39	47.6	951	2	Q6FK34_CANGA	Q6fk34 candida gla
433	39	47.6	442	2	Q37N81_RHOPA	Q37n81 rhodopseudo	506	39	47.6	1023	1	DPOL_ADEB3	Q3tp88 mus musculu
434	39	47.6	450	2	Q4IDP2_GIBZE	Q4idd2 gibberella	507	39	47.6	1023	1	DPOL_ADEB3	Q3tp88 mus musculu
435	39	47.6	460	2	Q6DX87_9INFA	Q6dx87 influenza a	508	39	47.6	1048	2	Q3J4A5_RHOSA	Q3j4a5 rhodobacter
436	39	47.6	461	2	Q6DMW3_9INFA	Q6dmw3 influenza a	509	39	47.6	1050	2	Q4X0U9_ASPFU	Q4x0u9 aspergillus
437	39	47.6	462	2	Q6DMW2_9INFA	Q6dmw2 influenza a	510	39	47.6	1097	2	Q870T4_NEUCR	Q870t4 neurospora
438	39	47.6	462	2	Q5DL12_9INFA	Q5dl12 influenza a	511	39	47.6	1098	2	Q43VM3_SOLUS	Q43vm3 solibacter
439	39	47.6	466	2	Q6DMW1_9INFA	Q6dmw1 influenza a	512	39	47.6	1124	2	Q5KBB6_CRYNE	Q5kbb6 cryptococcu
440	39	47.6	466	2	Q6DMW7_9INFA	Q6dmw7 influenza a	513	39	47.6	1134	2	Q6IRN3_XENLA	Q6irn3 xenopus lae
441	39	47.6	466	2	Q6DMW7_9INFA	Q6dmw7 influenza a	514	39	47.6	1189	2	Q2VB19_CHICK	Q2vb19 gallus gall
442	39	47.6	467	2	Q6DX55_9INFA	Q6dx55 influenza a	515	39	47.6	1190	2	Q3R784_XYLFA	Q3r784 xyella fas
443	39	47.6	469	2	Q98077_9INFA	Q98077 influenza a	516	39	47.6	1190	2	Q87DV4_XYLFT	Q87dv4 xyella fas
444	39	47.6	470	1	NRAM_IADM2	Q07573 influenza a	517	39	47.6	1209	2	Q9PDP9_XYLFA	Q9pdp9 xyella fas
445	39	47.6	470	1	NRAM_IAMGD	Q07577 influenza a	518	39	47.6	1209	2	Q3R1M8_XYLFA	Q3r1m8 xyella fas
446	39	47.6	470	1	NRAM_IAMAE	Q07583 influenza a	519	39	47.6	1209	2	Q3RH13_XYLFA	Q3rh13 xyella fas
447	39	47.6	470	1	NRAM_IAMKL	Q07584 influenza a	520	39	47.6	1210	2	Q85MY6_CRYNE	Q85my6 cryptococcu
448	39	47.6	470	2	Q59522_IATKC	Q07585 influenza a	521	39	47.6	1241	2	Q3TUE2_MOUSE	Q3tue2 mus musculu
449	39	47.6	470	2	Q5UG07_9INFA	Q59522 influenza a	522	39	47.6	1372	1	COLA2_MOUSE	Q01149 mus musculu
450	39	47.6	470	2	Q6DMW1_9INFA	Q6dmw1 influenza a	523	39	47.6	1372	1	COLA2_MOUSE	Q01149 mus musculu
451	39	47.6	470	2	Q6DMW9_9INFA	Q6dmw9 influenza a	524	39	47.6	1372	2	Q3TU64_MOUSE	Q3tu64 mus musculu
452	39	47.6	470	2	Q6DX39_9INFA	Q6dx39 influenza a	525	39	47.6	1372	2	Q3TXS7_MOUSE	Q3txs7 m osteoclas
453	39	47.6	470	2	Q6DX39_9INFA	Q6dx39 influenza a	526	39	47.6	1381	2	Q8TRM0_METAC	Q8trm0 methanosarc
454	39	47.6	470	2	Q6DXE3_9INFA	Q6dxe3 influenza a	527	39	47.6	1512	2	Q3WB2_9GAMM	Q3wb2 shewanella
455	39	47.6	470	2	Q7TF27_9INFA	Q7tf27 influenza a	528	39	47.6	1538	2	Q9SMG9_ARATH	Q9smg9 arabidopsis
456	39	47.6	472	2	Q6DXF9_9INFA	Q6dxf9 influenza a	529	39	47.6	1539	2	Q5NTX0_ADICAN	Q5ntx0 adiantum ca
457	39	47.6	484	2	Q3PJA2_PARDE	Q3pia2 paracoccus	530	39	47.6	1732	2	Q4LDF5_CHICK	Q4ldf5 gallus gall
458	39	47.6	500	2	Q445U0_SOLUS	Q445j0 solibacter	531	39	47.6	1734	2	Q89LV5_BRAJA	Q89lv5 bradyrhizob
459	39	47.6	506	2	Q37Q70_SFPHR	Q37q70 novosphingo	532	39	47.6	1737	2	Q6N6R1_RHOPA	Q6n6r1 rhodopseudo
460	39	47.6	519	2	Q4BTX7_BURVI	Q4btx7 burkholderi	533	39	47.6	1738	2	Q3PWE1_NITHA	Q3pwe1 nitrobacter
461	39	47.6	526	1	COLA2_RABIT	Q28668 oryctolagus	534	39	47.6	1743	2	Q3SF44_9BRAD	Q3sf44 bradyrhizob
462	39	47.6	528	2	Q4AY15_9BURK	Q4ay15 polaromonas	535	39	47.6	1760	2	Q9SHK6_ARATH	Q9shk6 arabidopsis
463	39	47.6	537	2	Q43S03_SOLUS	Q43s03 solibacter	536	39	47.6	4151	2	Q7D7J7_MYCTU	Q7d7j7 mycobacteri
464	39	47.6	543	2	Q5R4N2_PONPY	Q5r4n2 pongo pygma	537	39	47.6	4151	2	Q53490_MYCTU	Q53490 mycobacteri
465	39	47.6	544	2	Q2UTF7_ASPOR	Q2utf7 aspergillus	538	39	47.6	4151	2	Q7VES2_MYCBO	Q7ves2 mycobacteri
466	39	47.6	544	2	Q36QEA_MARHY	Q36qea marinobacte	539	38.5	47.0	99	2	Q6B6Z6_RABIT	Q6b6z6 oryctolagus
467	39	47.6	547	2	Q4B0Y9_9BURK	Q4b0y9 polaromonas	540	38.5	47.0	183	2	Q6XXU0_9VIRU	Q6xxu0 bamboo mosa
468	39	47.6	564	2	Q89J47_BRAJA	Q89j47 bradyrhizob	541	38.5	47.0	183	2	Q6XXU7_9VIRU	Q6xxu7 bamboo mosa
469	39	47.6	566	2	Q91VL4_MOUSE	Q91vl4 mus musculu	542	38.5	47.0	220	2	Q3BYZ5_XANC5	Q3byz5 xanthomonas

543	38.5	47.0	314	2	Q743T7_MYCPA	Q743T7	mycobacteri	616	38	46.3	244	2	Q6Q414_LYCES	O6Q414	lycopersico
544	38.5	47.0	382	2	Q2X991_PSEPU	Q2X991	pseudomonas	617	38	46.3	246	2	Q2J143_ARATH	O2J143	arabidopsis
545	38.5	47.0	382	2	Q8C69_PSEPK	Q8C69	pseudomonas	618	38	46.3	248	1	AP23_ARATH	P42736	arabidopsis
546	38.5	47.0	442	2	Q7TF28_9INFA	Q7TF28	influenza a	619	38	46.3	248	1	PT16_LYCES	O04682	lycopersico
547	38.5	47.0	561	2	Q5Q727_CRYNE	Q5Q727	cryptococcus	620	38	46.3	249	2	Q67MY7_SYMTH	O67MY7	symbiobacte
548	38.5	47.0	561	2	Q5KFY9_CRYNE	Q5KFY9	cryptococcus	621	38	46.3	251	2	O6S242_TOBAC	O6S242	nicotiana t
549	38.5	47.0	735	2	Q5C8D8_9THEM	Q5C8D8	thermotoga	622	38	46.3	252	2	O2HY03_PENCA	O2HY03	penicillium
550	38	46.3	84	2	Q98N19_RHILO	Q98N19	rhizobium l	623	38	46.3	254	2	Q479V4_DECAR	Q479V4	dechloromon
551	38	46.3	87	2	Q44933_9POTV	Q44933	plum pox vi	624	38	46.3	258	2	Q332T4_RHOPA	Q332T4	rhodopsendo
552	38	46.3	99	2	Q48BV2_PSE14	Q48BV2	pseudomonas	625	38	46.3	260	2	Q84XB2_LYCES	Q84XB2	lycopersico
553	38	46.3	101	2	Q6PDS4_MOUSE	Q6PDS4	mus musculus	626	38	46.3	260	2	Q6RJ36_LYCES	Q6RJ36	lycopersico
554	38	46.3	102	2	Q8WY95_HUMAN	Q8WY95	homo sapien	627	38	46.3	263	2	Q92NLO_RHIME	Q92NLO	rhizobium m
555	38	46.3	106	2	Q95M02_HORSE	Q95M02	equus caball	628	38	46.3	264	2	Q6Q8S5_CAPAN	Q6Q8S5	capsicum an
556	38	46.3	118	2	Q6K8B4_ORYSA	Q6K8B4	oryza sativ	629	38	46.3	270	2	Q3W647_9ACTO	Q3W647	frankia sp.
557	38	46.3	124	2	Q7XDZ2_ORYSA	Q7XDZ2	oryza sativ	630	38	46.3	272	2	Q32W72_CAPAN	Q32W72	capsicum an
558	38	46.3	128	2	Q6J988_ARATH	Q6J988	arabidopsis	631	38	46.3	272	2	Q2S1U8_9SPHI	Q2S1U8	salinibacte
559	38	46.3	131	2	Q480S2_COLP3	Q480S2	colwellia p	632	38	46.3	274	2	Q5DZRI_VIBF1	Q5DZRI	vibrio fisc
560	38	46.3	133	2	Q455G9_9BURK	Q455G9	burkholderi	633	38	46.3	275	1	ECX2_SULTO	Q97509	sulfolobus
561	38	46.3	133	2	Q4L1G1_9BURK	Q4L1G1	burkholderi	634	38	46.3	276	2	Q4JB28_SULAC	Q4JB28	sulfolobus
562	38	46.3	136	2	Q8X8X1_STRCO	Q8X8X1	streptomyce	635	38	46.3	276	2	Q32W74_CAPAN	Q32W74	capsicum an
563	38	46.3	139	2	Q9LTC6_ARATH	Q9LTC6	arabidopsis	636	38	46.3	276	2	Q9CY15_MOUSE	Q9CY15	mus musculu
564	38	46.3	141	2	Q6H7R2_ORYSA	Q6H7R2	oryza sativ	637	38	46.3	277	2	Q394Q7_BURS3	Q394Q7	burkholderi
565	38	46.3	141	2	Q8H075_ORYSA	Q8H075	oryza sativ	638	38	46.3	279	2	Q2IKK1_9DELT	Q2IKK1	anaeromyxob
566	38	46.3	144	2	Q96NC1_HUMAN	Q96NC1	homo sapien	639	38	46.3	280	2	O81365_PRUAR	O81365	prunus arme
567	38	46.3	144	2	Q4T6Z2_TETNG	Q4T6Z2	tetradodon n	640	38	46.3	282	1	ERF6_ARATH	Q8V231	arabidopsis
568	38	46.3	146	2	Q8LI66_ORYSA	Q8LI66	oryza sativ	641	38	46.3	289	2	Q37QC5_SPHAR	Q37QC5	novosphingo
569	38	46.3	147	2	Q5AUL9_EMENI	Q5AUL9	aspergillus	642	38	46.3	292	2	Q5IR87_HUMAN	Q5IR87	homo sapien
570	38	46.3	154	2	Q95M03_HORSE	Q95M03	equus caball	643	38	46.3	292	2	Q67RC7_SYMTH	Q67RC7	symbiobacte
571	38	46.3	156	2	Q4J756_SULAC	Q4J756	sulfolobus	644	38	46.3	293	2	Q4ASD3_9BURK	Q4ASD3	polaromonas
572	38	46.3	162	2	Q5VMU7_ORYSA	Q5VMU7	oryza sativ	645	38	46.3	297	2	Q4NFR1_9MICC	Q4NFR1	arthrobacte
573	38	46.3	162	2	Q8SB09_ORYSA	Q8SB09	oryza sativ	646	38	46.3	297	2	Q66012_BORBU	Q66012	borrelia ga
574	38	46.3	165	2	Q75UJ5_CUCME	Q75UJ5	cucumis mel	647	38	46.3	297	2	O51631_BORBU	O51631	borrelia bu
575	38	46.3	171	2	Q3XUN0_9PROT	Q3XUN0	magnetococc	648	38	46.3	298	2	Q2SK64_9GAMM	Q2SK64	hahella che
576	38	46.3	171	2	Q84B14_VIBF1	Q84B14	vibrio fisc	649	38	46.3	300	2	Q33MB5_METHU	Q33MB5	methanospir
577	38	46.3	172	2	Q7X1L1_ORYSA	Q7X1L1	oryza sativ	650	38	46.3	302	2	Q6W1U6_RHISN	Q6W1U6	rhizobium s
578	38	46.3	173	2	Q5Y4C5_ARATH	Q5Y4C5	arabidopsis	651	38	46.3	302	2	Q5Z214_NOCPA	Q5Z214	nocardia fa
579	38	46.3	182	2	Q3AX61_SYNS9	Q3AX61	synechococc	652	38	46.3	309	2	Q7BU48_STRAM	Q7BU48	streptomyce
580	38	46.3	183	2	Q7BTW3_YERPE	Q7BTW3	yersinia pe	653	38	46.3	314	2	Q231B5_SULTO	Q231B5	sulfolobus
581	38	46.3	183	2	Q2KWZ7_BORAV	Q2KWZ7	borderella	654	38	46.3	317	2	O973B3_SULTO	O973B3	sulfolobus
582	38	46.3	185	2	Q7R0F2_GIALA	Q7R0F2	giardia lam	655	38	46.3	319	2	O672K6_GOSBA	O672K6	gossypium b
583	38	46.3	188	2	Q82QNA_STRAW	Q82QNA	streptomyce	656	38	46.3	321	2	Q619N9_CABER	Q619N9	caenorhabdi
584	38	46.3	190	2	Q3GA65_LYCES	Q3GA65	lycopersico	657	38	46.3	322	2	Q8BLU3_MOUSE	Q8BLU3	mus musculu
585	38	46.3	190	2	Q6AMA7_DESPS	Q6AMA7	desulfotale	658	38	46.3	323	2	Q97UL6_SULSO	Q97UL6	sulfolobus
586	38	46.3	192	2	Q4ELK3_TRYCR	Q4ELK3	trypanosoma	659	38	46.3	324	2	Q4IFS5_GIBZE	Q4IFS5	gibberella
587	38	46.3	192	2	Q3G5V8_9DELT	Q3G5V8	pelobacter	660	38	46.3	327	2	O2NB19_9SPHN	O2NB19	erythrobact
588	38	46.3	193	2	Q8S8T1_ARATH	Q8S8T1	arabidopsis	661	38	46.3	329	2	O6QX54_COCPA	O6QX54	coffea cane
589	38	46.3	194	2	Q6YVX0_ORYSA	Q6YVX0	oryza sativ	662	38	46.3	330	2	Q6S5V3_9POTV	Q6S5V3	plum pox vi
590	38	46.3	195	2	Q68725_YERPE	Q68725	yersinia pe	663	38	46.3	332	2	Q6E5V2_9POTV	Q6E5V2	plum pox vi
591	38	46.3	195	2	Q7ARG3_YERPE	Q7ARG3	yersinia pe	664	38	46.3	333	2	Q6E5V1_9POTV	Q6E5V1	plum pox vi
592	38	46.3	195	2	Q935A9_SALT1	Q935A9	salmonella	665	38	46.3	335	1	DRE2A_ARATH	O82132	arabidopsis
593	38	46.3	196	2	Q3CRC8_ALTAT	Q3CRC8	pseudalter	666	38	46.3	335	2	O693B3_ORYSA	O693B3	oryza sativ
594	38	46.3	201	2	Q53N79_ORYSA	Q53N79	oryza sativ	667	38	46.3	335	2	O6QJK9_9LAMI	O6QJK9	fraxinus pe
595	38	46.3	207	1	FRDA_MOUSE	Q35943	mus musculu	668	38	46.3	338	2	Q4D5F6_TRYCR	Q4D5F6	trypanosoma
596	38	46.3	207	2	Q9KLJ4_VIBCH	Q9KLJ4	vibrio chol	669	38	46.3	338	2	Q2LDQ6_ARAHY	Q2LDQ6	arachis hyp
597	38	46.3	207	2	Q3TV21_MOUSE	Q3TV21	mus musculu	670	38	46.3	340	2	O6Q9W4_9BRAS	O6Q9W4	thellungiel
598	38	46.3	207	2	Q3UG34_MOUSE	Q3UG34	mus musculu	671	38	46.3	340	2	O8PPG0_XANAC	O8PPG0	xanthomonas
599	38	46.3	210	1	FRDA_HUMAN	Q16595	homo sapien	672	38	46.3	340	2	O6PW64_9INFA	O6PW64	influenza a
600	38	46.3	210	1	FRDA_MACPA	Q8bxx9	macaca fasc	673	38	46.3	340	2	Q6PW67_9INFA	Q6PW67	influenza a
601	38	46.3	210	2	Q5VZ01_HUMAN	Q5VZ01	homo sapien	674	38	46.3	342	2	Q7X826_ORYSA	Q7X826	oryza sativ
602	38	46.3	210	2	Q9LENE_CATRO	Q9LENE	catharanthu	675	38	46.3	344	2	Q8GTES_CICAR	Q8GTES	cicer ariet
603	38	46.3	211	2	Q3YRL5_HRCJ	Q3YRL5	ehrllichia c	676	38	46.3	345	2	Q3SE63_9BRAD	Q3SE63	bradyrhizob
604	38	46.3	211	2	Q5HAS8_EHRRW	Q5HAS8	ehrllichia r	677	38	46.3	346	2	O8JZ32_ARMV	O8JZ32	arabis mosa
605	38	46.3	214	2	Q3FR33_LYCES	Q3FR33	lycopersico	678	38	46.3	358	2	Q8L9V3_ARATH	Q8L9V3	arabidopsis
606	38	46.3	217	2	Q5FD64_EHRRW	Q5FD64	ehrllichia r	679	38	46.3	358	2	Q8SSA8_ARATH	Q8SSA8	arabidopsis
607	38	46.3	217	2	Q5FFU6_EHRRG	Q5FFU6	ehrllichia r	680	38	46.3	362	1	QUEA_DEIRA	Q9RUI9	deinococcus
608	38	46.3	218	1	ERF1B_ARATH	Q8ldc8	arabidopsis	681	38	46.3	362	2	O69Y47_ORYSA	O69Y47	oryza sativ
609	38	46.3	218	2	Q5YN71_NOCPA	Q5YN71	nocardia fa	682	38	46.3	362	2	Q9SP16_ORYSA	Q9SP16	oryza sativ
610	38	46.3	222	2	Q9AP70_9BACT	Q9AP70	uncultured	683	38	46.3	366	2	Q2JU07_9CYAN	Q2JU07	cyanobacter
611	38	46.3	222	2	Q9AP71_9BACT	Q9AP71	uncultured	684	38	46.3	374	1	RM02_YARLI	Q6C9C4	yarrowia li
612	38	46.3	234	2	Q2LUS9_9DELT	Q2LUS9	syntrophus	685	38	46.3	375	2	Q382R2_9TRYP	Q382R2	trypanosoma
613	38	46.3	235	2	Q3HW33_CAPAN	Q3HW33	capsicum an	686	38	46.3	375	2	Q8VXK9_FAGSY	Q8VXK9	fagus sylv
614	38	46.3	239	2	Q8H312_ORYSA	Q8H312	oryza sativ	687	38	46.3	378	1	YN9B_YEAST	P52923	saccharomyc
615	38	46.3	239	2	Q93A09_THIFE	Q93A09	thiobacillu	688	38	46.3	378	2	Q70AB2_FAGSY	Q70AB2	fagus sylv

689	38	46.3	380	2	Q5ICH9_9ROSI	Q5icn9	populus alb	762	38	46.3	469	2	Q6XV56_9INFA	Q6xv56	influenza a
690	38	46.3	381	2	Q52QX1_WANES	Q52qy1	manihot esc	763	38	46.3	469	2	Q6XV57_9INFA	Q6xv57	influenza a
691	38	46.3	385	2	Q2PEZ6_TRIPR	Q2pez6	trifolium p	764	38	46.3	469	2	Q6XV59_9INFA	Q6xv59	influenza a
692	38	46.3	387	2	Q7Y1W2_TOBAC	Q7y1w2	nicotiana t	765	38	46.3	469	2	Q6XV60_9INFA	Q6xv60	influenza a
693	38	46.3	389	2	Q37G65_RHOPA	Q37g65	rhodopseu	766	38	46.3	469	2	Q6XV64_9INFA	Q6xv64	influenza a
694	38	46.3	390	2	Q3L0R1_GOSHI	Q3l0r1	gossypium h	767	38	46.3	469	2	Q6XV65_9INFA	Q6xv65	influenza a
695	38	46.3	390	2	Q34C06_RHOPA	Q34c06	rhodopseu	768	38	46.3	469	2	Q6XV66_9INFA	Q6xv66	influenza a
696	38	46.3	390	2	Q3PS76_NITHA	Q3ps76	nitrobacter	769	38	46.3	469	2	Q6XV69_9INFA	Q6xv69	influenza a
697	38	46.3	390	2	Q3SNC2_NITWN	Q3snc2	nitrobacter	770	38	46.3	469	2	Q6XV70_9INFA	Q6xv70	influenza a
698	38	46.3	390	2	Q2J016_RHOPA	Q2j016	rhodopseu	771	38	46.3	469	2	Q6XV72_9INFA	Q6xv72	influenza a
699	38	46.3	390	2	Q6N1U5_RHOPA	Q6n1u5	rhodopseu	772	38	46.3	469	2	Q7TF24_9INFA	Q7tf24	influenza a
700	38	46.3	392	2	Q4JDR2_9PROT	Q4jdr2	nitrosospir	773	38	46.3	469	2	Q7TF23_9INFA	Q7tf23	influenza a
701	38	46.3	394	2	Q2RPH0_RHORU	Q2rph0	rhodospiril	774	38	46.3	469	2	Q80M09_9INFA	Q80mj9	influenza a
702	38	46.3	396	2	Q3KZ43_GOSHI	Q3kz43	gossypium h	775	38	46.3	469	2	Q82S55_9INFA	Q82s55	influenza a
703	38	46.3	399	2	Q4HQZ6_CAMUP	Q4hqz6	campylobact	776	38	46.3	469	2	Q8B197_9INFA	Q8b197	influenza a
704	38	46.3	402	1	MURD_CAMJR	Q5hw34	campylobact	777	38	46.3	469	2	Q8B198_9INFA	Q8b198	influenza a
705	38	46.3	405	2	Q39LQ6_BURS3	Q39lq6	burkholderi	778	38	46.3	469	2	Q8JSE0_9INFA	Q8jse0	influenza a
706	38	46.3	416	1	CCA_SULAC	Q4j9a0	sulfolobus	779	38	46.3	469	2	Q91B11_9INFA	Q91bi1	influenza a
707	38	46.3	417	2	Q6BJW5_DEBHA	Q6bjw5	debaromyce	780	38	46.3	469	2	Q91B12_9INFA	Q91bi2	influenza a
708	38	46.3	418	2	Q93H82_STRAW	Q93h82	streptomyce	781	38	46.3	469	2	Q9WJT1_9INFA	Q9wjt1	influenza a
709	38	46.3	424	2	Q4I530_GIBZE	Q4i530	gibberella	782	38	46.3	469	2	Q9VNL3_9INFA	Q9vnl3	influenza a
710	38	46.3	425	2	Q4WB80_ASPFU	Q4wb80	aspergillus	783	38	46.3	469	2	Q9IGQ0_IANJ	Q9igq0	influenza a
711	38	46.3	427	2	Q2LFT2_9INFA	Q2lft2	influenza a	784	38	46.3	469	2	Q2LFS5_9INFA	Q2lfs5	influenza a
712	38	46.3	430	2	Q6FUF6_CANGA	Q6fuf6	candida gla	785	38	46.3	469	2	Q2LFS7_9INFA	Q2lfs7	influenza a
713	38	46.3	432	2	Q483Y5_COLP3	Q483y5	colwellia p	786	38	46.3	469	2	Q693C5_9INFA	Q693c5	influenza a
714	38	46.3	437	1	WTM1_YEAST	Q12363	saccharomyc	787	38	46.3	469	2	Q6DQJ4_9INFA	Q6dqj4	influenza a
715	38	46.3	437	2	Q5AQ89_EMENI	Q5aq89	aspergillus	788	38	46.3	469	2	Q6DSV3_9INFA	Q6dsv3	influenza a
716	38	46.3	437	2	Q3RT53_RALME	Q3rt53	raistonia m	789	38	46.3	469	2	Q6LZB9_9INFA	Q6lzb9	influenza a
717	38	46.3	441	2	Q6XV63_9INFA	Q6xv63	influenza a	790	38	46.3	469	2	Q89473_9INFA	Q89473	influenza a
718	38	46.3	442	2	Q2LFT3_9INFA	Q2lft3	influenza a	791	38	46.3	469	2	Q21DG4_9INFA	Q21dga	influenza a
719	38	46.3	443	2	Q4WYV3_ASPFU	Q4wyv3	aspergillus	792	38	46.3	470	1	NRAM_IACHU	Q7570	influenza a
720	38	46.3	443	2	Q7TF26_9INFA	Q7tf26	influenza a	793	38	46.3	470	1	NRAM_IADBU	Q7571	influenza a
721	38	46.3	444	2	Q9I107_PSEAE	Q9i107	pseudomonas	794	38	46.3	470	1	NRAM_IADCH	Q7572	influenza a
722	38	46.3	445	2	Q7TF25_9INFA	Q7tf25	influenza a	795	38	46.3	470	1	NRAM_IADH2	Q7579	influenza a
723	38	46.3	445	2	Q2LFT4_9INFA	Q2lft4	influenza a	796	38	46.3	470	1	NRAM_IADU3	Q7576	influenza a
724	38	46.3	446	2	Q6GYV3_9INFA	Q6gyv3	influenza a	797	38	46.3	470	1	NRAM_IAHAI	Q7578	influenza a
725	38	46.3	446	2	Q6GYV4_9INFA	Q6gyv4	influenza a	798	38	46.3	470	1	NRAM_IAHUJ	Q7578	influenza a
726	38	46.3	446	2	Q701T6_9INFA	Q701t6	influenza a	799	38	46.3	470	1	NRAM_IAKTB	Q7578	influenza a
727	38	46.3	446	2	Q701U1_9INFA	Q701u1	influenza a	800	38	46.3	470	1	NRAM_IALEN	Q7578	influenza a
728	38	46.3	447	2	Q8H0Y2_ARATH	Q8h0y2	arabidopsis	801	38	46.3	470	2	Q595Z8_9INFA	Q595z8	influenza a
729	38	46.3	447	2	Q9FI18_ARATH	Q9fi18	arabidopsis	802	38	46.3	470	2	Q5EDD2_9INFA	Q5edd2	influenza a
730	38	46.3	449	2	Q4FB58_9INFA	Q4fb58	influenza a	803	38	46.3	470	2	Q5WMA3_9INFA	Q5wma3	influenza a
731	38	46.3	449	2	Q5Q028_9INFA	Q5q028	influenza a	804	38	46.3	470	2	Q67218_9INFA	Q67218	influenza a
732	38	46.3	449	2	Q2L9W4_9INFA	Q2l9w4	influenza a	805	38	46.3	470	2	Q67219_9INFA	Q67219	influenza a
733	38	46.3	453	1	NRAM_IANL	P03470	influenza a	806	38	46.3	470	2	Q67220_9INFA	Q67220	influenza a
734	38	46.3	454	1	NRAM_IAPUE	P03468	influenza a	807	38	46.3	470	2	Q67221_9INFA	Q67221	influenza a
735	38	46.3	454	2	Q6DX47_9INFA	Q6dx47	influenza a	808	38	46.3	470	2	Q67222_9INFA	Q67222	influenza a
736	38	46.3	454	2	Q8JU04_9INFA	Q8juu4	influenza a	809	38	46.3	470	2	Q7TG84_9INFA	Q7tg84	influenza a
737	38	46.3	455	2	Q94E25_ARATH	Q94ez5	arabidopsis	810	38	46.3	470	2	Q7TG89_9INFA	Q7tg89	influenza a
738	38	46.3	455	2	Q2L2N6_9INFA	Q2l2n6	influenza a	811	38	46.3	470	2	Q7TG90_9INFA	Q7tg90	influenza a
739	38	46.3	460	2	Q8V3X8_9INFA	Q8v3x8	influenza a	812	38	46.3	470	2	Q7TG91_9INFA	Q7tg91	influenza a
740	38	46.3	460	2	Q8V3X9_9INFA	Q8v3x9	influenza a	813	38	46.3	470	2	Q8B199_9INFA	Q8b199	influenza a
741	38	46.3	462	1	FUCO_RAT	P17164	rattus norv	814	38	46.3	470	2	Q8JMH6_9INFA	Q8jmh6	influenza a
742	38	46.3	462	2	Q2XK49_PSEPU	Q2xk49	pseudomonas	815	38	46.3	470	2	Q8JMH7_9INFA	Q8jmh7	influenza a
743	38	46.3	463	2	Q8V3X5_9INFA	Q8v3x5	influenza a	816	38	46.3	470	2	Q8JSD8_9INFA	Q8jstd8	influenza a
744	38	46.3	464	2	Q8M9B9_PPOAL	Q8m9b9	mesanthemum	817	38	46.3	470	2	Q9WA95_9INFA	Q9wa95	influenza a
745	38	46.3	466	2	Q5W4S0_SERRE	Q5w4s0	serenoa rep	818	38	46.3	470	2	Q2L1N42_9INFA	Q2l1n42	influenza a
746	38	46.3	466	2	Q2LFS6_9INFA	Q2lfs6	influenza a	819	38	46.3	470	2	Q2JHB8_9INFA	Q2jhb8	influenza a
747	38	46.3	468	2	Q5IR88_HUMAN	Q5ir88	homo sapien	820	38	46.3	475	2	Q4J622_SULAC	Q4j622	sulfolobus
748	38	46.3	468	2	Q4H2D1_9INFA	Q4h2d1	influenza a	821	38	46.3	475	2	Q8M9C1_9POAL	Q8m9c1	mesanthemum
749	38	46.3	469	2	Q6DX95_9INFA	Q6dx95	influenza a	822	38	46.3	477	2	Q8ID62_PLAF7	Q8id62	plasmodium
750	38	46.3	469	2	Q6GYU9_9INFA	Q6gyu9	influenza a	823	38	46.3	488	2	Q9MRA0_9CARY	Q9mra0	silene nuta
751	38	46.3	469	2	Q6GYU9_9INFA	Q6gyu9	influenza a	824	38	46.3	489	2	Q9WY7_9ROSI	Q9wty7	salix retic
752	38	46.3	469	2	Q6GYV0_9INFA	Q6gyv0	influenza a	825	38	46.3	493	2	Q915R4_PSEAE	Q915r4	pseudomonas
753	38	46.3	469	2	Q6GYV5_9INFA	Q6gyv5	influenza a	826	38	46.3	496	2	Q9BA92_TRAFO	Q9ba92	trachycarpu
754	38	46.3	469	2	Q6GYV6_9INFA	Q6gyv6	influenza a	827	38	46.3	498	2	Q7HHY7_LICGR	Q7hh7	licuala gra
755	38	46.3	469	2	Q6GYV7_9INFA	Q6gyv7	influenza a	828	38	46.3	498	2	Q9B127_9LILI	Q9b127	pritchardio
756	38	46.3	469	2	Q6GYV8_9INFA	Q6gyv8	influenza a	829	38	46.3	498	2	Q9BA90_9LILI	Q9ba90	livistona s
757	38	46.3	469	2	Q6GYV9_9INFA	Q6gyv9	influenza a	830	38	46.3	498	2	Q9BA94_9LILI	Q9ba94	rhapis subt
758	38	46.3	469	2	Q6XV31_9INFA	Q6xv31	influenza a	831	38	46.3	498	2	Q9BA96_CHAHU	Q9ba96	chamaerops
759	38	46.3	469	2	Q6XV37_9INFA	Q6xv37	influenza a	832	38	46.3	502	2	Q445X3_SOLUS	Q445x3	solibaerops
760	38	46.3	469	2	Q6XV54_9INFA	Q6xv54	influenza a	833	38	46.3	514	2	Q800L5_ANGJA	Q800l5	anguilla ja
761	38	46.3	469	2	Q6XV55_9INFA	Q6xv55	influenza a	834	38	46.3	516	2	Q342T2_RHOPA	Q342t2	rhodopseu



835	38	46.3	516	2	Q447R5 SOLUS	Q447R5 solibacter	908	38	46.3	1032	2	Q8ZV80 PYRAE	Q8ZV80 pyrobaculum
836	38	46.3	520	2	Q2T769 BURTH	Q2t769 burkholderi	909	38	46.3	1032	2	Q7JLZ2 CAEBL	Q7j1z2 caenorhabdi
837	38	46.3	520	2	P72647 SYNY3	P72647 synechocyst	910	38	46.3	1032	2	Q61J44 CAEBR	Q61j44 caenorhabdi
838	38	46.3	529	2	Q6ZJEB ORYSA	Q6zje2 oryza sativ	911	38	46.3	1035	2	Q21079 CAEBL	Q21079 caenorhabdi
839	38	46.3	531	2	Q7WZ63 PACTO	Q7wz63 nonomuraea	912	38	46.3	1040	1	Y043 CAEBL	F34681 caenorhabdi
840	38	46.3	538	2	Q443S5 SOLUS	Q443s5 solibacter	913	38	46.3	1053	2	Q3VVZ6 PROAE	Q3vvz6 prosthecoch
841	38	46.3	541	2	Q340T4 RHOPA	Q340t4 rhodopseudo	914	38	46.3	1054	2	Q5R5N8 PONGY	Q5r5n8 pongo pygma
842	38	46.3	544	2	Q6N6G0 RHOPA	Q6n6g0 rhodopseudo	915	38	46.3	1073	2	Q69Z56 MOUSE	Q69z56 mus musculu
843	38	46.3	554	2	Q89LD3 BRAJA	Q89ld3 bradyrhizob	916	38	46.3	1097	2	Q9HGT2 CANAL	Q9hgt2 candida alb
844	38	46.3	556	2	Q45233 BRAJA	Q45232 bradyrhizob	917	38	46.3	1114	2	Q4SXG7 TETNG	Q4sxn7 tetradodon n
845	38	46.3	560	2	Q4HXW1 GIBZE	Q4hxn1 gibberella	918	38	46.3	1117	2	Q5IRH9 HUMAN	Q5irh9 homo sapien
846	38	46.3	560	2	Q73NKO TREDE	Q73nko treponema d	919	38	46.3	1141	2	Q2WQJ7 HUMAN	Q2wqj7 homo sapien
847	38	46.3	567	2	Q8KCI0 CHLTE	Q8kci0 chlorobium	920	38	46.3	1187	2	Q5EMX8 9POTV	Q5emx8 plum pox vi
848	38	46.3	570	2	Q354Q9 9BRAD	Q354q9 bradyrhizob	921	38	46.3	1209	2	Q7XA27 SOLBU	Q7xa27 solanum bul
849	38	46.3	576	2	Q46Q20 RALEU	Q46q20 ralstonia e	922	38	46.3	1226	2	Q4J5B1 AZOVI	Q4j5b1 azotobacter
850	38	46.3	578	2	Q60J20 CAEBR	Q60j20 caenorhabdi	923	38	46.3	1307	2	Q49AJ0 HUMAN	Q49aj0 homo sapien
851	38	46.3	579	2	Q82P17 STRAW	Q82p17 streptomyce	924	38	46.3	1358	2	Q6FVM9 CANGA	Q6fvm9 candida gla
852	38	46.3	585	2	Q3FQQ3 9BURK	Q3fqq3 rhodofera	925	38	46.3	1366	1	CO1A2 CANFA	Q46392 canis famil
853	38	46.3	591	2	Q9ZG10 THIFE	Q9zgi10 thioabacillu	926	38	46.3	1426	2	Q4Q868 LEIMA	Q4q868 leishmania
854	38	46.3	597	2	Q3B6T8 ARATH	Q3b6t8 arabidopsis	927	38	46.3	1453	1	CO1A1 CHICK	P02457 gallus gall
855	38	46.3	597	2	Q5M7M1 XENTR	Q5m7m1 xenopus tro	928	38	46.3	1469	2	Q36VQ6 RHOPA	Q36vq6 rhodopseudo
856	38	46.3	601	2	Q3QMD6 9RHOB	Q3qmd6 silicibacte	929	38	46.3	1565	2	Q54CE3 DICDI	Q54ce3 dictyosteli
857	38	46.3	602	2	Q4P939 USTMA	Q4p939 ustilago ma	930	38	46.3	1737	2	Q37E12 RHOPA	Q37e12 rhodopseudo
858	38	46.3	613	2	Q96794 DROME	Q96794 drosophila	931	38	46.3	1739	2	Q3SSE5 NITWN	Q3sse5 nitrobacter
859	38	46.3	613	2	Q9VJG0 DROME	Q9vjg0 drosophila	932	38	46.3	1867	2	Q4RI01 TETNG	Q4ri01 tetradodon n
860	38	46.3	621	2	Q84FAB MYAXA	Q84fab myxococcu	933	38	46.3	3096	2	Q4GXV0 9POTV	Q4gxv0 lily motle
861	38	46.3	623	2	Q3H8C6 TRIER	Q3h8c6 trichodesmi	934	38	46.3	3140	1	POLG PPVPA	P17767 p genome po
862	38	46.3	625	2	Q826C5 STRAW	Q826c5 streptomyce	935	38	46.3	3140	1	POLG PPVSK	Q84934 p genome po
863	38	46.3	627	2	Q54G06 DICDI	Q54g06 dictyosteli	936	38	46.3	3140	2	Q2VQ18 9POTV	Q2vq18 plum pox vi
864	38	46.3	635	2	Q3PM99 NITHA	Q3pm99 nitrobacter	937	38	46.3	3140	2	Q2VQ19 9POTV	Q2vq19 plum pox vi
865	38	46.3	637	2	Q3EE13 ACTSC	Q3eel3 actinobacil	938	38	46.3	3140	2	Q2VQ20 9POTV	Q2vq20 plum pox vi
866	38	46.3	644	2	Q9ECM7 HUMAN	Q9ecm7 homo sapien	939	38	46.3	3140	2	Q2VQ21 9POTV	Q2vq21 plum pox vi
867	38	46.3	659	2	Q3DY01 CHLAU	Q3dy01 chloroflexu	940	38	46.3	3140	2	Q2VQ22 9POTV	Q2vq22 plum pox vi
868	38	46.3	661	2	Q2V3C6 ARATH	Q2v3c6 arabidopsis	941	38	46.3	3140	2	Q2VQ23 9POTV	Q2vq23 plum pox vi
869	38	46.3	662	1	KAT3 ARATH	P92960 arabidopsis	942	38	46.3	3140	2	Q2VQ24 9POTV	Q2vq24 plum pox vi
870	38	46.3	663	2	Q52WC5 BACLD	Q62wcs bacillus li	943	38	46.3	3140	2	Q52V18 9POTV	Q52v18 plum pox vi
871	38	46.3	665	2	Q55XK8 BACLD	Q65kx8 bacillus li	944	38	46.3	3140	2	Q52V19 9POTV	Q52v19 plum pox vi
872	38	46.3	683	2	Q3YP51 NOCFA	Q5yp51 nocardia fa	945	38	46.3	3140	2	Q80S19 9POTV	Q80s19 plum pox vi
873	38	46.3	688	2	Q4CBJ3 CLOTM	Q4cbj3 clostridium	946	38	46.3	3140	2	Q80S20 9POTV	Q80s20 plum pox vi
874	38	46.3	689	2	Q2RLK6 MOOTH	Q2rlk6 moorella th	947	38	46.3	3140	2	Q84925 9POTV	Q84925 plum pox vi
875	38	46.3	696	2	Q6FVL3 CANGA	Q6fvl3 candida gla	948	38	46.3	3140	2	Q98WJ5 9POTV	Q98wj5 plum pox vi
876	38	46.3	705	2	Q6X4V5 GOSHI	Q6x4v5 gossypium h	949	38	46.3	3140	2	Q9PYF0 9POTV	Q9pyf0 plum pox vi
877	38	46.3	744	2	Q4RLD6 TETNG	Q4rlde tetradodon n	950	38	46.3	3141	2	Q52V21 9POTV	Q52v21 plum pox vi
878	38	46.3	747	1	PN1 YARLI	Q6cee7 yarrowia li	951	38	46.3	3587	2	Q6RKK6 GIBMO	Q6rkk6 gibberella
879	38	46.3	750	2	Q8N1N8 HUMAN	Q8n1n8 homo sapien	952	38	46.3	4684	2	Q7R1S4 GIALA	Q7r1s4 giardia lam
880	38	46.3	753	2	Q5YU9 NOCFA	Q5yu9 nocardia fa	953	38	46.3	6939	2	Q2T1K7 BURTH	Q2t1k7 burkholderi
881	38	46.3	760	2	Q4AIR7 9CHLB	Q4air7 chlorobium	954	37.5	45.7	108	2	Q2T1K7 BURTH	Q2t1k7 burkholderi
882	38	46.3	763	2	Q4IKV5 METBU	Q4ikv5 methanococc	955	37.5	45.7	236	2	Q3B0E5 SYN9S	Q3b0e5 synechococc
883	38	46.3	799	2	Q82KF5 STRAW	Q82kf5 streptomyce	956	37.5	45.7	244	2	Q4H5U6 9DEIO	Q4h5u6 deinocococc
884	38	46.3	810	1	NELL1 RAT	Q62919 rattus norv	957	37.5	45.7	256	2	Q7U9S0 SYNXP	Q7u9s0 synechococ
885	38	46.3	810	2	Q2VWQ2 MOUSE	Q2vwq2 mus musculu	958	37.5	45.7	279	2	Q8U7B4 AGRT5	Q8u7b4 agrobacteri
886	38	46.3	822	2	Q5D1Z2 9POTV	Q5d1z2 plum pox vi	959	37.5	45.7	282	2	Q7NU47 CHRVO	Q7nu47 chromobacte
887	38	46.3	824	2	Q7VLM9 HAEDU	Q7vlm9 haemophilus	960	37.5	45.7	290	2	Q9P827 CANAL	Q9p827 candida alb
888	38	46.3	858	2	Q6FN07 CANGA	Q6fn07 candida gla	961	37.5	45.7	297	2	Q59RM2 CANAL	Q59rm2 candida alb
889	38	46.3	860	1	AT56 HUMAN	Q9ukp5 homo sapien	962	37.5	45.7	297	2	Q59S23 CANAL	Q59s23 candida alb
890	38	46.3	860	2	Q6D8U4 ERWCT	Q6d8u4 erwinia car	963	37.5	45.7	304	2	Q7CV72 AGRT5	Q7cv72 agrobacteri
891	38	46.3	866	2	Q4U4W7 THEAU	Q4u4w7 thermoaescu	964	37.5	45.7	338	2	Q353P1 9GAMW	Q353p1 alkalilimni
892	38	46.3	907	2	Q988S0 CARAU	Q988s0 carassius a	965	37.5	45.7	345	2	Q7MEZ0 VIBVU	Q7mez0 vibrio vuln
893	38	46.3	913	2	Q2XEL2 PSEPU	Q2xel2 pseudomonas	966	37.5	45.7	345	2	Q8D7X3 VIBVU	Q8d7x3 vibrio vuln
894	38	46.3	916	2	Q5BA51 EMENI	Q5ba51 aspergillus	967	37.5	45.7	360	2	Q2RF44 RHODPR	Q2rf44 rhodospirill
895	38	46.3	929	2	Q7F8S7 ORYSA	Q7f8s7 oryza sativ	968	37.5	45.7	382	2	Q48BQ8 PSE14	Q48bq8 pseudomonas
896	38	46.3	947	2	Q8TF46 HUMAN	Q8tf46 homo sapien	969	37.5	45.7	382	2	Q87U42 PSESM	Q87u42 pseudomonas
897	38	46.3	955	2	Q3B6W2 PELLD	Q3b6w2 pelodictyon	970	37.5	45.7	436	2	Q8BWH5 MOUSE	Q8bwh5 mus musculu
898	38	46.3	970	2	Q8C0S1 MOUSE	Q8c0s1 mus musculu	971	37.5	45.7	440	2	Q7V3Y1 PROMM	Q7v3y1 prochloroco
899	38	46.3	971	2	Q8WTU9 HUMAN	Q8wtu9 homo sapien	972	37.5	45.7	507	2	Q4W907 ASPFU	Q4w907 aspergillus
900	38	46.3	971	2	Q5U2P0 RAT	Q5u2p0 rattus norv	973	37.5	45.7	574	2	Q84FK7 ENTAG	Q84fk7 enterobacte
901	38	46.3	976	2	Q73SX6 MYCPA	Q73sx6 mycobacteri	974	37.5	45.7	582	2	Q4DV56 TRYCR	Q4dv56 trypanosoma
902	38	46.3	983	2	Q5ZL00 CHICK	Q5z100 gallus gall	975	37.5	45.7	679	2	Q2U8S2 ASPOR	Q2u8s2 aspergillus
903	38	46.3	994	2	Q2JGC9 9ACTO	Q2jgc9 frankia sp.	976	37.5	45.7	734	2	Q5CBP9 THENE	Q5cbp9 thermotoga
904	38	46.3	999	2	Q3C207 9VIRU	Q3c207 heterocapsa	977	37.5	45.7	735	2	Q5CBG4 9THEM	Q5cbg4 thermotoga
905	38	46.3	1003	2	Q39S98 GROMG	Q3c207 heterocapsa	978	37.5	45.7	747	2	Q4WIU2 ASPFU	Q4wiu2 aspergillus
906	38	46.3	1004	2	Q3C205 9VIRU	Q3c205 heterocapsa	979	37.5	45.7	772	2	Q89PP4 BRAJA	Q89pp4 bradyrhizob
907	38	46.3	1010	1	CLPP CHLEU	P43279 chlamydomon	980	37.5	45.7	1004	1	YD83 SCHPO	Q10408 schizosacch

981 37.5 45.7 1401 2 Q80YP9 mus musculus  
 982 37.5 45.7 2581 2 Q8ZS63 anabaena sp  
 983 37.5 45.7 2766 2 Q3VZE3 frankia sp.  
 984 37 45.1 64 2 Q96GX8 homo sapien  
 985 37 45.1 100 2 Q4WHB8 aspergillus  
 986 37 45.1 101 2 Q3HJ08 ASPFU  
 987 37 45.1 101 2 Q3HJ08 TRIER  
 988 37 45.1 101 2 Q4C709 CROWT  
 989 37 45.1 106 2 Q4KL44 MOUSE  
 990 37 45.1 108 2 Q3FHN0 9BURK  
 991 37 45.1 118 2 Q62WM3 BACLD  
 992 37 45.1 119 2 Q6Z355 ORYSA  
 993 37 45.1 119 2 Q7UJ28 RHOB  
 994 37 45.1 121 2 Q6G3F9 BIFLO  
 995 37 45.1 129 2 Q6D326 ERWCA  
 996 37 45.1 131 2 Q651A5 ORYSA  
 997 37 45.1 131 2 Q6J9R1 ARATH  
 998 37 45.1 131 2 Q9LSX0 ARATH  
 999 37 45.1 131 2 Q43U69 SOLUS  
 1000 37 45.1 132 2 Q87S66 VIBPA  
 1000 37 45.1 136 2 Q4B963 BURVI

## ALIGNMENTS

## RESULT 1

HPSE HUMAN  
 ID HPSE\_HUMAN STANDARD; PRT; 543 AA.  
 AC Q3Y25L; Q53GE5; Q9UL39;  
 DT 11-NOV-2005, integrated into UniProtKB/Swiss-Prot.  
 DT 01-NOV-1999, sequence version 1.  
 DT 07-FEB-2006, entry version 27.  
 DE Heparanase precursor [EC 3.2.-.-] (Heparanase-1) (Hpal) (Endo-  
 DE glucuronidase) [Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa  
 DE subunit].  
 GN Name=HPSE; Synonyms=HEP, HPA, HPA1, HPRI, HPSE1, HSE1;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.  
 RC TISSUE=Placenta;  
 RX MEDLINE=9935379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;  
 RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C.,  
 RA Seddon A.P., Giorgio N.A., Bohlens P.;  
 RT "Cloning and functional expression of a human heparanase gene";  
 RL Biochem. Biophys. Res. Commun. 261:183-187(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN  
 RP SEQUENCE OF 158-168; 326-337 AND 447-491.  
 RC TISSUE=Embryonic fibroblast;  
 RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;  
 RA Toyoshima M., Nakajima M.;  
 RT "Human heparanase. Purification, characterization, cloning, and  
 RT expression.";  
 RL J. Biol. Chem. 274:24153-24160(1999).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.  
 RX PubMed=10395325; DOI=10.1038/10518;  
 RA Vlodavsky I., Friedmann Y., Elkin M., Aingorn H., Atzmon R.,  
 RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,  
 RA Spector L., Pecker I.;  
 RT "Mammalian heparanase: gene cloning, expression and function in tumor  
 RT progression and metastasis";  
 RL Nat. Med. 5:793-802(1999).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE  
 RP OF 158-174; 263-272; 326-337; 433-436; 438-443; 466-468 AND 478-483.  
 RC TISSUE=Placenta;  
 RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;  
 RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,

Parish C.R.;  
 RT "Cloning of mammalian heparanase, an important enzyme in tumor  
 RT invasion and metastasis";  
 RL Nat. Med. 5:803-809(1999).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE  
 RP SPECIFICITY.  
 RC TISSUE=Placenta;  
 RX MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;  
 RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;  
 RT "Heparanase expression in invasive trophoblasts and acute vascular  
 RT damage";  
 RL Glycobiology 10:467-475(2000).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.  
 RX PubMed=11547900; DOI=10.1023/A:1011375624902;  
 RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O.,  
 RA Aviv A., Elkin M., Pecker I., Peretz T., Vlodavsky I.;  
 RT "Molecular properties and involvement of heparanase in cancer  
 RT progression and mammary gland morphogenesis";  
 RL J. Mammary Gland Biol. Neoplasia 6:311-322(2001).  
 RN [7]  
 RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163,  
 RP SUBUNITS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.  
 RC TISSUE=Placenta;  
 RX PubMed=12713442; DOI=10.1042/BJ20030318;  
 RA McKenzie E., Young K., Hircok M., Bennett J., Bhanan M., Felix R.,  
 RA Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S.,  
 RA Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M.,  
 RA Parekh R., Stubberfield C.;  
 RT "Biochemical characterization of the active heterodimer form of human  
 RT heparanase (Hpal) protein expressed in insect cells";  
 RL Biochem. J. 373:423-435(2003).  
 RN [8]  
 RP NUCLEOTIDE SEQUENCE [MRNA].  
 RP Pinnal M.A., Semedo P.;  
 RT "Cloned heparanase from MCF-7 cells";  
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.  
 RN [9]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
 RC TISSUE=Small intestine;  
 RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,  
 RA Tanaka A., Yokoyama S.;  
 RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.  
 RN [10]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
 RC TISSUE=Pancreas;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [11]  
 RP MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.  
 RX PubMed=11123890; DOI=10.1021/bi002080p;  
 RA Hulett M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,  
 RA Greedy J.E., Parish C.R.;

RT "Identification of active-site residues of the pro-metastatic  
RL endoglycosidase heparanase.";  
RN Biochemistry 39:15659-15667(2000).  
RN [12]  
RP N-GLYCOSYLATION, AND MUTAGENESIS OF ASN-162; ASN-178; ASN-200;  
RP ASN-217; ASN-238 AND ASN-459.  
RX PubMed:14573609; DOI=10.1074/jbc.M300541200;  
RA Simizu S., Ishida K., Wierzbz M.K., Osada H.;  
RT Secretion of heparanase protein is regulated by glycosylation in  
RT human tumor cell lines.";  
RL J. Biol. Chem. 279:2697-2703(2004).  
RN [13]  
RN SUBCELLULAR LOCATION.  
RX PubMed:15292202; DOI=10.1074/jbc.M402131200;  
RA Gingis-Velitski S., Zetser A., Kaplan V., Ben-Zaken O., Cohen E.,  
RA Levy-Adam F., Bashenko Y., Flugelman M.Y., Vlodavsky I., Ilan N.;  
RT "Heparanase uptake is mediated by cell membrane heparan sulfate  
RT proteoglycans.";  
RL J. Biol. Chem. 279:44084-44092(2004).  
RN [14]  
RN SUBCELLULAR LOCATION, PROCESSING, AND MUTAGENESIS OF TYR-156.  
RX PubMed:15659389; DOI=10.1074/jbc.M413370200;  
RA Aboud-Jarrous G., Rangini-Guetta Z., Aingorn H., Atzmon R.,  
RA Elgavish S., Peretz T., Vlodavsky I.;  
RT "Site-directed mutagenesis, proteolytic cleavage, and activation of  
RT human proheparanase.";  
RL J. Biol. Chem. 280:13568-13575(2005).  
RN [16]  
RN DOMAINS, AND MUTAGENESIS OF LYS-158 AND LYS-161.  
RX PubMed:15760902; DOI=10.1074/jbc.M414546200;  
RA Levy-Adam F., Aboud-Jarrous G., Guerrini M., Beccati D.,  
RA Vlodavsky I., Ilan N.;  
RT "Identification and characterization of heparin/heparan sulfate  
RT binding domains of the endoglycosidase heparanase.";  
RL J. Biol. Chem. 280:20457-20466(2005).  
RN [17]  
RN VARIANT SER-260.  
RX PubMed:15334672;  
RA Chen X.P., Liu Y.B., Rui J., Peng S.Y., Peng C.H., Zhou Z.Y.,  
RA Shi L.H., Shen H.W., Xu B.;  
RT "Heparanase mRNA expression and point mutation in hepatocellular  
RT carcinoma.";  
RL World J. Gastroenterol. 10:2795-2799(2004).  
CC -1- FUNCTION: Endoglycosidase which is a cell surface and  
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
CC proteoglycans. Also implicated in the extravasation of leukocytes  
CC and tumor cell lines. Due to its contribution to metastasis and  
CC angiogenesis, it is considered to be a potential target for anti-  
CC cancer therapies.  
CC -1- ENZYME REGULATION: Inhibited by EDTA, laminarin sulfate and, to a  
CC lower extent, by heparin and sulfamin and activated by calcium and  
CC magnesium (By similarity).  
CC -1- BIOPHYSICOCHEMICAL PROPERTIES:  
CC Optimum pH is 4-6;  
CC SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
CC subunits, the proteolytic products.  
CC -1- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.  
CC Secreted, internalised and transferred to late endosomes/lysosomes  
CC as a proheparanase. In lysosomes, it is processed into the active  
CC form, the heparanase. The uptake or internalisation of  
CC proheparanase is mediated by HSPGs. Heparin appears to be a  
CC competitor and retain proheparanase in the extracellular medium.  
CC -1- TISSUE SPECIFICITY: Highly expressed in placenta and spleen and  
CC weakly expressed in lymph node, thymus, peripheral blood  
CC leukocytes, bone marrow, endothelial cells, fetal liver and tumor

CC tissues.  
CC -1- PTM: Proteolytically processed. The cleavage of the 65 kDa form  
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa  
CC product. The active form, the 8/50 kDa heterodimer, is resistant  
CC to degradation. Complete removal of the linker peptide appears to  
CC be a prerequisite to the complete activation of the enzyme.  
CC -1- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears  
CC to be essential for its solubility.  
Query Match 100.0%; Score 82; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 3.8e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RPQKVKWLGETSSAY 15  
DB 334 RPQKVKWLGETSSAY 348  
RESULT 2  
Q33X5 SPAJD PRELIMINARY; PRT; 558 AA.  
AC Q33X5;  
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.  
DT 06-DEC-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Heparanase.  
GN Name=hpa;  
OS Spalax judaei (Blind subterranean mole rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Spalacidae; Spalacinae; Spalax.  
OX NCBI\_TaxID=134510;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
RT cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
RT cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AM085494; CAJ30021.1; -; mRNA.  
SQ SEQUENCE 558 AA; 62737 MW; 07BAF8F55849BEE7 CRC64;  
Query Match 100.0%; Score 82; DB 2; Length 558;  
Best Local Similarity 100.0%; Pred. No. 3.9e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RPQKVKWLGETSSAY 15  
DB 349 RPQKVKWLGETSSAY 363  
RESULT 3  
Q33X6 SPAJD PRELIMINARY; PRT; 574 AA.  
AC Q33X6;  
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.  
DT 06-DEC-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Heparanase.  
GN Name=hpa;  
OS Spalax judaei (Blind subterranean mole rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



DR EMBL; AM085490; CAJ30017.1; -; mRNA.  
 SQ SEQUENCE 574 AA; 64525 MW; 1635865051B380D0 CRC64;  
 Query Match 100.0%; Score 82; DB 2; Length 574;  
 Best Local Similarity 100.0%; Pred. No. 4e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPQKKVWLGTSAY 15  
 |||||  
 DB 365 RPQKKVWLGTSAY 379  
 RESULT 7  
 HPSE MOUSE  
 ID HPSE MOUSE STANDARD; PRT; 535 AA.  
 AC Q6VGGZ1; Q8K3K3;  
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
 DT 11-OCT-2005, sequence version 2.  
 DT 07-MAR-2006, entry version 13.  
 DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) [Contains:  
 DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].  
 GN Name=Hpse; Synonyms=Hpa;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridea; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA].  
 RC STRAIN=SJL/J; TISSUE=Spleen;  
 RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;  
 RA Hulet M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,  
 RA Parish C.R.;  
 RT "Cloning of mammalian heparanase, an important enzyme in tumor  
 RT invasion and metastasis";  
 RL Nat. Med. 5:803-809(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 28-57 AND 150-179,  
 RP GLYCOSYLATION, BIOPHYSICO-CHEMICAL PROPERTIES, ENZYME REGULATION, AND  
 RP SUBUNITS.  
 RC STRAIN=FVB; TISSUE=Embryo;  
 RX MEDLINE=22350326; PubMed=12460766; DOI=10.1016/S1046-5928(02)00558-2;  
 RA Maio H.-Q., Navarro E., Patel S., Sargent D., Koo H., Wan H.,  
 RA Plata A., Zhou Q., Ludwig D., Bohlen P., Kussie P.;  
 RT "Cloning, expression, and purification of mouse heparanase";  
 RL Protein Expr. Purif. 26:425-431(2002).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.  
 RX MEDLINE=22841152; PubMed=12837765; DOI=10.1074/jbc.M300925200;  
 RA Gong F., Jemth P., Galvis M.L.E., Vlodavsky I., Horner A., Lindahl U.,  
 RA Li J.-P.;  
 RT "Processing of macromolecular heparin by heparanase";  
 RL J. Biol. Chem. 278:35152-35158(2003).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
 RC STRAIN=C57BL/6J, and NOD; TISSUE=Thymus;  
 RX PubMed=16141072; DOI=10.1126/science.112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
 RA Ambesi-Impombato A., Apweiler R., Attaliva R.N., Bailey T.L.,  
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Hill D., Humnienick L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
 RA Jakt M., Kanapin A., Katoh M., Kawasaki Y., Kelso J., Kitamura H.,  
 RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,

Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 Matsuura H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 Mortagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,  
 Nilsson Y., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavoni G., Pesole G.,  
 Retovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
 Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
 Schonbach C., Sekiguchi K., Sempke C.A., Seno S., Sees L., Sheng Y.,  
 Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,  
 Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
 Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
 Nishio T., Okada M., Plesky C., Shibata K., Shiraki T., Suzuki S.,  
 Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 Hayashizaki Y.;  
 RA "The transcriptional landscape of the mammalian genome";  
 RT Science 309:1559-1563(2005).  
 RL  
 CC -!- FUNCTION: Endoglycosidase which is a cell surface and  
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
 CC proteoglycans. Also implicated in the extravasation of leukocytes  
 CC and tumor cell lines. Contributes to metastasis and angiogenesis  
 CC (By similarity).  
 CC -!- ENZYME REGULATION: Inhibited by EDTA and activated by calcium and  
 CC magnesium (By similarity). Inhibited by laminarin sulfate and, to  
 CC a lower extent, by heparin and sulfamin.  
 CC -!- BIOPHYSICO-CHEMICAL PROPERTIES:  
 CC pH dependence:  
 CC Optimum pH is 5;  
 CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
 CC subunits, the proteolytic products.  
 CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.  
 CC secreted, internalised and transferred to late endosomes/lysosomes  
 CC as a proheparanase. In lysosomes, it is processed into the active  
 CC form, the heparanase. The uptake or internalisation of  
 CC proheparanase is mediated by HSPGs. Heparin appears to be a  
 CC competitor and retain proheparanase in the extracellular medium  
 CC (By similarity).  
 CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form  
 CC leads to the generation of a linker peptide, 8 kDa and 50 kDa  
 CC product. The active form, the 8/50 kDa heterodimer, is resistant  
 CC to degradation. Complete removal of the linker peptide appears to  
 CC be a prerequisite to the complete activation of the enzyme (By  
 CC similarity).  
 CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears  
 CC to be essential for its solubility.  
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 CC EMBL; AF359507; AAQ15188.1; -; mRNA.  
 CC EMBL; AY077467; AAL76083.1; -; mRNA.  
 CC EMBL; AY151051; AAN41636.1; -; mRNA.  
 CC EMBL; AK040471; BAC30600.1; -; mRNA.  
 CC EMBL; AK154628; BAE32725.1; -; mRNA.  
 CC Ensembl; ENSMUSG0000035273; Mus musculus.  
 CC MGI; MGI:1343124; Hpsae.  
 CC GO; GO:0005578; C:extracellular matrix (sensu Metazoa); TAS.  
 CC InterPro; IPR005199; Glyco\_hydro\_79\_N.  
 CC Pfam; PF03662; Glyco\_hydro\_79n; 1.  
 CC Calcium; Direct protein sequencing; Glycoprotein; Hydrolase; Lysosome;  
 CC Magnesium; Membrane; Signal.  
 CC SIGNAL 1 27 By similarity.  
 CC CHAIN 28 101 Heparanase 8 kDa subunit.  
 CC /FTId=PRO\_0000042263.  
 CC PROPEP 102 149 Linker peptide (By similarity).  
 CC

```
FT CHAIN 150 535 /FTid=PRO_0000042264.
FT Heparanase 50 kDa subunit.
FT /FTid=PRO_0000042265.
FT Heparin/HS-binding (By similarity).
FT Heparin/HS-binding (By similarity).
FT Proton donor (Potential).
FT Nucleophile (Potential).
FT N-linked (GlcNAc. . .) (Potential).
FT N-linked (GlcNAc. . .) (Potential).
FT N-linked (GlcNAc. . .) (Potential).
FT N-linked (GlcNAc. . .) (Potential).
FT N-linked (GlcNAc. . .) (Potential).
FT N-linked (GlcNAc. . .) (Potential).
FT K -> R (in Ref. 3).
FT CONFLICT 206 206 W -> S (in Ref. 3).
FT CONFLICT 212 212 NGS -> DGL (in Ref. 1, 2 and 4).
FT CONFLICT 230 232 E -> K (in Ref. 3).
FT CONFLICT 335 335 G -> A (in Ref. 3).
FT CONFLICT 342 342 Y -> H (in Ref. 1, 2 and 4).
FT CONFLICT 455 455 V -> I (in Ref. 1, 2 and 4).
FT CONFLICT 531 531
SQ SEQUENCE 535 AA; 60050 MW; AF19B28B7CD03F7B CRC64;

Query Match 93.9%; Score 77; DB 1; Length 535;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSSAY 15
DB 327 PGKKVWLGETSSAY 340

RESULT 8
HPSE RAT
ID HPSE RAT STANDARD; PRT; 536 AA.
AC Q9I9FL; Q9QZP8;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 05-JUL-2004, sequence version 1.
DT 07-MAR-2006, entry version 11.
DE Heparanase precursor [EC 3.2.-.-] (Endo-glucuronidase) [Contains:
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].
DE Name=Hpse; Synonyms=Hep;
GN Rattus norvegicus (Rat).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
OX [1]
RN NUCLEOTIDE SEQUENCE [MRNA].
RP MEDLINE=22194309; PubMed=10395326; DOI=10.1038/10525;
RX Podyma-Inoue K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;
RT "Characterization of heparanase from a rat parathyroid cell line.";
RL J. Biol. Chem. 277:32459-32465 (2002).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Contributes to metastasis and angiogenesis
CC (By similarity).
CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower
CC extent, by heparin and sulfamin (By similarity). Activated by
CC calcium and magnesium. Inhibited by EDTA.
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products (By similarity).
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes
```

```
CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium
CC (By similarity).
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme (By
CC similarity).
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AF359508; AAQ15189.1; -; mRNA.
CC EMBL; AF184967; AAF04563.1; -; mRNA.
DR RGD; 61969; Hpse.
DR InterPro: IPR005199; Glyco_hydro_79_N.
DR Pfam; PF03662; Glyco_hydro_79n; 1.
KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;
KW Signal.
FT SIGNAL 1 28 By similarity.
FT CHAIN 29 102 Heparanase 8 kDa subunit.
FT PROPEP 103 150 /FTid=PRO_0000042266.
FT CHAIN 151 536 Linker peptide (By similarity).
FT /FTid=PRO_0000042267.
FT Heparanase 50 kDa subunit.
FT /FTid=PRO_0000042268.
FT REGION 151 155 Heparin/HS-binding (By similarity).
FT ACT_SITE 263 273 Heparin/HS-binding (By similarity).
FT ACT_SITE 218 218 Proton donor (Potential).
FT CARBOHYD 336 336 Nucleophile (Potential).
FT CARBOHYD 155 155 N-linked (GlcNAc. . .) (By similarity).
FT CARBOHYD 193 193 N-linked (GlcNAc. . .) (By similarity).
FT CARBOHYD 210 210 N-linked (GlcNAc. . .) (By similarity).
FT CARBOHYD 452 452 N-linked (GlcNAc. . .) (By similarity).
FT CONFLICT 15 15 G -> R (in Ref. 2).
FT CONFLICT 227 227 H -> Q (in Ref. 2).
FT CONFLICT 350 350 D -> N (in Ref. 2).
SQ SEQUENCE 536 AA; 60480 MW; C434E04CF536EA4D CRC64;

Query Match 93.9%; Score 77; DB 1; Length 536;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSSAY 15
DB 328 PGKKVWLGETSSAY 341

RESULT 9
HPSE CHICK
ID HPSE CHICK STANDARD; PRT; 523 AA.
AC Q90TK5;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-DEC-2001, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Heparanase precursor (EC 3.2.-.-).
DE Name=HPSE; Synonyms=HPA;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
OX [1]
RN NUCLEOTIDE SEQUENCE [MRNA].
RP MEDLINE=21369959; PubMed=11387326; DOI=10.1074/jbc.M102462200;
RX Goldshmidt O., Zcharia E., Aingorn H., Guatta-Rangini Z., Atzmon R.,
RA Michal I., Pecker I., Mitrani E., Vlodavsky I.;
```



"Expression pattern and secretion of human and chicken heparanase are determined by their signal peptide sequence.";  
 J. Biol. Chem. 276:29178-29187(2001).  
 CC -!- FUNCTION: Endoglycosidase which is a cell surface and extracellular matrix-degrading enzyme. Cleaves heparan sulfate proteoglycans (HSPGs) into heparan sulfate side chains and core proteoglycans (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted (By similarity).  
 CC -!- PTM: N-glycosylated (By similarity).  
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL; AY037007; AAK82648.1; -; mRNA.  
 DR EMBL; ENSGALG00000011203; Gallus gallus.  
 DR InterPro; IPR005199; Glyco\_hydro\_79\_N.  
 DR Pfam; PF03662; Glyco\_hydro\_79n; 1.  
 KW Glycoprotein; Hydrolase; Lysosome; Membrane; Signal.  
 FT SIGNAL 1 18 Potential.  
 FT CHAIN 19 523 Heparanase.  
 FT REGION 137 141 /FTid=PRO\_0000042259.  
 FT REGION 250 260 Heparin/HS-binding (By similarity).  
 FT ACT\_SITE 204 204 Heparin/HS-binding (By similarity).  
 FT ACT\_SITE 323 323 Proton donor (Potential).  
 FT CARBOHYD 141 141 Nucleophile (Potential).  
 FT CARBOHYD 141 141 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 196 196 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 436 436 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 439 439 N-linked (GlcNAc...) (Potential).  
 SQ SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF881 CRC64;  
 Query Match 89.0%; Score 73; DB 1; Length 523;  
 Best Local Similarity 92.9%; Pred. No. 0.0013;  
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 PGKKVWLGETSSAY 15  
 DB 315 PGKKVWLGETGSAY 328  
 RESULT 10  
 Q4SYF6 TETNG PRELIMINARY; PRT; 533 AA.  
 AC Q4SYF6;  
 DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
 DT 19-JUL-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 4.  
 DE Chromosome undetermined SCAF12073, whole genome shotgun sequence.  
 DE (Fragment).  
 DE ORFNames=GSTENG0010356001;  
 GN Tetraodon nigroviridis (Green puffer).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; OC Tetraodontidae; Tetraodontidae; Tetraodon.  
 OC NCBI\_TaxID=99883;  
 RN NUCLEOTIDE SEQUENCE.  
 RP PubMed=15496914; DOI=10.1038/nature03025;  
 RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N., RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A., RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B., RA DaSilva C., Salanoubat M., Levy M., Boudet N., Castellano S., RA Anchoard V., Jubin C., Castelli V., Katinka M., Vacherie B., RA Blemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V., RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J., RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S., RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J., RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M., RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C., RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.,

"Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype.";  
 Nature 431:946-957(2004).  
 RN NUCLEOTIDE SEQUENCE.  
 RP Genoscope; Whitehead Institute Centre for Genome Research; RG Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL; CAAE01012073; CAF94326.1; -; Genomic\_DNA.  
 DR EMBL; CAAE01012073; CAF94326.1; -; Genomic\_DNA.  
 FT NON\_TER 1 1  
 FT NON\_TER 533 533  
 SQ SEQUENCE 533 AA; 60100 MW; 9B00A7C8780100FF CRC64;  
 Query Match 86.6%; Score 71; DB 2; Length 533;  
 Best Local Similarity 92.9%; Pred. No. 0.0029;  
 Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 PGKKVWLGETSSAY 15  
 DB 291 PGKPVLGETSSAY 304  
 RESULT 11  
 HPSE\_BOVIN STANDARD; PRT; 545 AA.  
 ID HPSE\_BOVIN STANDARD; PRT; 545 AA.  
 AC Q9MYI0;  
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
 DT 01-JUN-2001, sequence version 2.  
 DT 07-MAR-2006, entry version 15.  
 DE Heparanase precursor (EC 3.2.-.-) [Contains: Heparanase 8 kDa subunit; DE Heparanase 50 kDa subunit].  
 GN Name=HPSE;  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Euthera; Laurasiatheria; Cetartiodactyla; Ruminantia; OC Pecora; Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RN NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.  
 RP TISSUE=Placenta;  
 RC MEDLINE=21176669; PubMed=11277877;  
 RA Kizaki K., Nakano H., Nakano H., Takahashi T., Imai K., Haeizume K.; RT "Expression of heparanase mRNA in bovine placenta during gestation.";  
 RL Reproduction 121:573-580(2001).  
 CC -!- FUNCTION: Endoglycosidase which is a cell surface and extracellular matrix-degrading enzyme. Cleaves heparan sulfate proteoglycans (HSPGs) into heparan sulfate side chains and core proteoglycans. Also implicated in the extravasation of leukocytes and tumor cell lines. Contributes to metastasis and angiogenesis (By similarity).  
 CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower extent, by heparin, sulfamin and EDTA. Activated by calcium and magnesium (By similarity).  
 CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa subunits, the proteolytic products (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted. Secreted, internalised and transferred to late endosomes/lysosomes as a proheparanase. In lysosomes, it is processed into the active form, the heparanase. The uptake or internalisation of proheparanase is mediated by HSPGs. Heparin appears to be a competitor and retain proheparanase in the extracellular medium (By similarity).  
 CC -!- TISSUE SPECIFICITY: Highly expressed in placenta and weakly in the kidney, lung, spleen and uterus.  
 CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form leads to the generation of a linker peptide, 8 kDa and 50 kDa product. The active form, the 8/50 kDa heterodimer, is resistant



CC to degradation. Complete removal of the linker peptide appears to  
 CC be a prerequisite to the complete activation of the enzyme (By  
 CC similarity).  
 CC -|- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears  
 CC to be essential for its solubility (By similarity).  
 CC -|- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL: AF281160; AAF87301.2; -; mRNA.  
 DR InterPro: IPR005199; Glyco\_hydro\_79\_N.  
 DR Pfam: PF03662; Glyco\_hydro\_79n; I.  
 DR Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;  
 KW Signal.  
 FT SIGNAL 1 37 By similarity.  
 FT CHAIN 38 111 Heparanase 8 kDa subunit (By similarity).  
 FT /FTid=PRO\_0000042256.  
 FT PROPEP 112 159 Linker peptide.  
 FT /FTid=PRO\_0000042257.  
 FT CHAIN 160 545 Heparanase 50 kDa subunit (By  
 FT similarity).  
 FT /FTid=PRO\_0000042258.  
 FT REGION 160 164 Heparin/HS-binding (Potential).  
 FT REGION 272 282 Heparin/HS-binding (Potential).  
 FT ACT SITE 227 227 Proton donor (Potential).  
 FT ACT SITE 345 345 Nucleophile (Potential).  
 FT CARBOHYD 164 164 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 219 219 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 461 461 N-linked (GlcNAc...) (Potential).  
 SQ SEQUENCE 545 AA; 61077 MW; FAC4BDFD855B933 CRC64;

Query Match 82.9%; Score 68; DB 1; Length 545;  
 Best Local Similarity 86.7%; Pred. No. 0.0099;  
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPKKKVWLGETSSAY 15  
 || |||||  
 DB 336 RPLKKVWLGETSSAF 350

RESULT 12  
 Q4TGC8\_TETNG  
 ID Q4TGC8\_TETNG PRELIMINARY; PRT; 255 AA.  
 AC Q4TGC8\_TETNG  
 DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
 DT 19-JUL-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 4.  
 DE Chromosome undetermined SCAF3783, whole genome shotgun sequence.  
 DE (Fragment).  
 DE ORFNames=GSTENG0001168001;  
 OS Tetraodon nigroviridis (Green puffer).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontidae; Tetraodontinae; Tetraodon.  
 OC NCBI\_TaxID=99883;  
 RN NUCLEOTIDE SEQUENCE.  
 RP PubMed=15496914; DOI=10.1038/nature03025;  
 RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,  
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,  
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,  
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
 RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,  
 RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,  
 RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Guzy J.,  
 RA Parra G., Lardier G., Chappell C., McKernan K.J., McEwan P., Bosak S.,  
 RA Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,  
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
 RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;  
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals  
 the early vertebrate proto-karyotype.";  
 RN Nature 431:946-957(2004).  
 RP NUCLEOTIDE SEQUENCE.  
 RG Genoscope; Whitehead Institute Centre for Genome Research;  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBSJ databases.  
 CC -|- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL: AF281160; AAF87301.2; -; Genomic\_DNA.  
 FT NON\_TER 1 255  
 SQ SEQUENCE 255 AA; 28562 MW; 07F542A9C755E3F0 CRC64;

RT the early vertebrate proto-karyotype.";  
 RL Nature 431:946-957(2004).  
 RP NUCLEOTIDE SEQUENCE.  
 RG Genoscope; Whitehead Institute Centre for Genome Research;  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBSJ databases.  
 CC -|- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL: CAAB01003783; CAF88054.1; -; Genomic\_DNA.  
 FT NON\_TER 1 255  
 SQ SEQUENCE 255 AA; 28562 MW; 07F542A9C755E3F0 CRC64;

Query Match 78.0%; Score 64; DB 2; Length 255;  
 Best Local Similarity 92.3%; Pred. No. 0.022;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 RPKKKVWLGETSSA 14  
 |||||  
 DB 171 RPKVWLGETSSA 183

RESULT 13  
 Q4TB80\_TETNG  
 ID Q4TB80\_TETNG PRELIMINARY; PRT; 597 AA.  
 AC Q4TB80;  
 DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
 DT 19-JUL-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 4.  
 DE Chromosome 17 SCAF7180, whole genome shotgun sequence. (Fragment).  
 DE ORFNames=GSTENG0003868001;  
 OS Tetraodon nigroviridis (Green puffer).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontidae; Tetraodontinae; Tetraodon.  
 OC NCBI\_TaxID=99883;  
 RN NUCLEOTIDE SEQUENCE.  
 RP PubMed=15496914; DOI=10.1038/nature03025;  
 RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,  
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,  
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,  
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
 RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,  
 RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,  
 RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Guzy J.,  
 RA Parra G., Lardier G., Chappell C., McKernan K.J., McEwan P., Bosak S.,  
 RA Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,  
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
 RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;  
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals  
 the early vertebrate proto-karyotype.";  
 RN Nature 431:946-957(2004).  
 RP NUCLEOTIDE SEQUENCE.  
 RG Genoscope; Whitehead Institute Centre for Genome Research;  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBSJ databases.  
 CC -|- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL: CAAB01007180; CAF89852.1; -; Genomic\_DNA.  
 FT NON\_TER 597

```
SQ SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;
Query Match 61.0%; Score 50; DB 2; Length 597;
Best Local Similarity 64.3%; Pred. No. 14;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 2 PGKKVWLGETSSAY 15
|||||||
|:
Db 375 PGKKVWLGLGPAW 388

RESULT 14
Q21PR2_9DELT PRELIMINARY; PRT; 586 AA.
AC Q21PR2_
DT 07-MAR-2006, integrated into UniProtKB/TrEMBL.
DT 07-MAR-2006, sequence version 1.
DT 07-MAR-2006, entry version 1.
DE Thiamine pyrophosphate enzyme.
GN ORFNames=Adel1015;
OS Anaeromyxobacter dehalogenans 2CP-C.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Cyobacterineae; Myxococcaceae; Anaeromyxobacter.
OX NCBI_TaxID=290397;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2CP-C;
RG US DOE Joint Genome Institute;
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Brettin T., Bruce D., Han C.,
RA Tapia R., Gilna P., Kiss H., Schmutz J., Larimer F., Land M.,
RA Kyrpides N., Anderson J., Sanford R.A., Ritalahti K.M., Thomas H.S.,
RA Kirby J.R., Zhulin I.B., Loeffler F.E., Richardson P.;
RA "Complete sequence of Anaeromyxobacter dehalogenans 2CP-C."
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000251; ABC80790.1; -; Genomic DNA.
SQ SEQUENCE 586 AA; 62020 MW; 39D77C7222A18EF7 CRC64;

Query Match 59.8%; Score 49; DB 2; Length 586;
Best Local Similarity 58.8%; Pred. No. 21;
Matches 10; Conservative 3; Mismatches 2; Indels 2; Gaps 1;
QY 1 RPKKVWL--GETSSAY 15
|||||||
|:
Db 466 RPGEVWLLYDGGSCAY 482

RESULT 15
Q36KN3_MARHY PRELIMINARY; PRT; 197 AA.
AC Q36KN3;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Phosphonate metabolism.
GN ORFNames=MaquidRAFT_4132;
OS Marinobacter aquaeolei VT8.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Marinobacter.
OX NCBI_TaxID=351348;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=VT8;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Richardson P.;
RA "Sequencing of the draft genome and assembly of Marinobacter aquaeolei VT8."
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.

SQ SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;
Query Match 61.0%; Score 50; DB 2; Length 597;
Best Local Similarity 64.3%; Pred. No. 14;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 2 PGKKVWLGETSSAY 15
|||||||
|:
Db 375 PGKKVWLGLGPAW 388

RESULT 16
Q3QDK4_9GAMW PRELIMINARY; PRT; 616 AA.
AC Q3QDK4;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Ferredoxin:Oxidoreductase FAD/NAD(P)-binding; MOSC; N-terminal
DE beta barrel; Oxidoreductase FAD-binding region.
GN ORFNames=SamadRAFT_1894;
OS Shewanella amazonensis SB2B.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=326297;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SB2B;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Richardson P.;
RA "Sequencing of the draft genome and assembly of Shewanella amazonensis SB2B."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAIN0100034; EAN37953.1; -; Genomic DNA.
SQ SEQUENCE 197 AA; 21261 MW; E86FF9188614ACC3 CRC64;

Query Match 58.5%; Score 48; DB 2; Length 197;
Best Local Similarity 61.5%; Pred. No. 9.7;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 2 PGKKVWLGETSSA 14
|||||||
|:
Db 42 PGESVWLADTDSA 54

RESULT 17
Q3QDK4_9GAMW PRELIMINARY; PRT; 616 AA.
AC Q3QDK4;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Ferredoxin:Oxidoreductase FAD/NAD(P)-binding; MOSC; N-terminal
DE beta barrel; Oxidoreductase FAD-binding region.
GN ORFNames=SamadRAFT_1894;
OS Shewanella amazonensis SB2B.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=326297;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SB2B;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RA "Annotation of the draft genome assembly of Shewanella amazonensis SB2B."
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAIN0100034; EAN37953.1; -; Genomic DNA.
SQ SEQUENCE 197 AA; 21261 MW; E86FF9188614ACC3 CRC64;

Query Match 58.5%; Score 48; DB 2; Length 197;
Best Local Similarity 61.5%; Pred. No. 9.7;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 2 PGKKVWLGETSSA 14
|||||||
|:
Db 42 PGESVWLADTDSA 54
```

```
DR InterPro; IPR005303; MOSC.N.
DR InterPro; IPR008333; Oxred FAD bd.
DR InterPro; IPR001433; Oxred FAD NAD bd.
DR InterPro; IPR001221; Phe hydroxylase.
DR Pfam; PF00970; FAD binding_6; 1.
DR Pfam; PF001111; Fer2; 1.
DR Pfam; PF03473; MOSC; 1.
DR Pfam; PF03476; MOSC.N; 1.
DR Pfam; PF00175; NAD binding 1; 1.
DR PRINTS; PR00410; PHEHYDRXLASE.
DR PROSITE; PS00197; 2FE2S_FER_1; 1.
DR PROSITE; PS1085; 2FE2S_FER_2; 1.
SQ SEQUENCE 616 AA; 67299 MW; EDB349A10D7494A4 CRC64;

Query Match 58.5%; Score 48; DB 2; Length 616;
Best Local Similarity 50.0%; Pred. No. 32;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSSAY 15
| : ||||| :
Db 122 PARLWLGETSNRF 135

RESULT 17
Q41J33 METBU PRELIMINARY; PRT; 334 AA.
AC Q41J33;
DT 27-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 27-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Trimethylamine methyltransferase.
GN OFRNAmec-MburaDRAFT_0457;
OS Methanococcoides burtonii DSM 6242.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanococcoides.
OX NCBI_TaxID=259564;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=DSM 6242;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Methanococcoides
RT burtonii DSM 6242.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
[2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=DSM 6242;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Methanococcoides burtonii
RT DSM 6242.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
[3]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=DSM 6242;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AADH02000014; EAM99398.1; -; Genomic DNA.
DR GO; GO:0008168; F:methyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0015948; P:methanogenesis; IEA.
DR InterPro; IPR010426; MTTB.
DR Pfam; PF06253; MTTB; 1.
```

```
KW Methyltransferase; Transferase.
SQ SEQUENCE 334 AA; 36229 MW; 5C8FCB81F0FE0D2A CRC64;

Query Match 57.3%; Score 47; DB 2; Length 334;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSSAY 15
| : ||||| :
Db 284 PGAKVWYGSSTAF 297

RESULT 18
Q54359 STRLN PRELIMINARY; PRT; 426 AA.
AC Q54359;
DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1996, sequence version 1.
DT 21-FEB-2006, entry version 25.
DE LmbF protein.
GN Name=lmbF;
OS Streptomyces lincolnensis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1915;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=78-11;
RX MEDLINE=96026646; PubMed=8577249;
RA Peschke U., Schmidt H., Zhang H.Z., Piepersberg W.;
RT "Molecular characterization of the lincomycin-production gene cluster
RT of Streptomyces lincolnensis 78-11.";
RL Mol. Microbiol. 16:1137-1156(1995).
CC -1- COPACTOR: Pyridoxal phosphate (By similarity).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; X79146; CAA55752.1; -; Genomic DNA.
DR PIR; S69815; S44953.
DR GO; GO:0016769; P:transferase activity, transferring nitrogen. .; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR004839; Aminotrans I/II.
DR Pfam; PF00155; Aminotran 1.2; 2-
DR Pyridoxal phosphate; Transferase.
KW SEQUENCE 426 AA; 46681 MW; 553250E88E083BA2 CRC64;

Query Match 57.3%; Score 47; DB 2; Length 426;
Best Local Similarity 60.0%; Pred. No. 33;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 RPQKKVWLGETSSAY 15
| : ||||| :
Db 395 RPWFKVWLGSDSVF 409

RESULT 19
Q6LAL3_9POTV PRELIMINARY; PRT; 3143 AA.
ID Q6LAL3_9POTV
AC Q6LAL3;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Polyprotein.
OS Plum pox virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
OC Potyvirus.
OX NCBI_TaxID=12211;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Sweet Cherry;
RX MEDLINE=98163291; PubMed=9504763; DOI=10.1016/S0166-0934(97)00158-4;
RX Crescenzi A., d'Aquino L., Nuzzaci M., Ostuni A., Bavoso A., Comes S.,
```

RA De Stradis A., Piazzolla P.;  
 RT "Production of strain specific antibodies against a synthetic  
 RT polypeptide corresponding to the N-terminal region of the plum pox  
 RT potyvirus coat protein."; J. Virol. Methods 69:181-189(1997).  
 RL [2]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=Sweet Cherry;  
 RC MEDLINE=22939292; PubMed=14579174; DOI=10.1007/s00705-003-0175-9;  
 RX Panigluolo A., Comes S., Maiss E., Piazzolla P., Crescenzi A.;  
 RA "The complete nucleotide sequence of Plum pox virus isolates from  
 RT sweet (PPV-SwC) and sour (PPV-Soc) cherry and their taxonomic  
 RT relationships within the species."; Arch. Virol. 148:2137-2153(2003).  
 RL Arch. Virol. 148:2137-2153(2003).  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 CC EMBL; Y09851; CAF02291.1; -; Genomic\_RNA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0006508; P:proteolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR011545; DEAD/DEAH\_N.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR002540; Pept\_S30\_Pot\_P1.  
 DR InterPro; IPR001730; Peptidase\_C4.  
 DR InterPro; IPR001456; Peptidase\_C6.  
 DR InterPro; IPR001592; Poty\_coat.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR001205; RNA\_pol\_P3D.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF00270; DEAD; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00863; Peptidase\_C4; 1.  
 DR Pfam; PF00851; Peptidase\_C6; 1.  
 DR Pfam; PF01577; Peptidase\_S30; 1.  
 DR Pfam; PF00767; Poty\_coat; 1.  
 DR Pfam; PF00680; RdRP\_1; 1.  
 DR PRINTS; PR00966; NIAPOTYPTASE.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELICC; 1.  
 KW Capsid protein; Polyprotein.  
 FT CHAIN 1 308 P1 protein.  
 FT CHAIN 309 766 HC protein.  
 FT CHAIN 767 1116 P3 protein.  
 FT CHAIN 1117 1168 6K1 protein.  
 FT CHAIN 1169 1803 CI protein.  
 FT CHAIN 1804 1856 6K2 protein.  
 FT CHAIN 1857 2293 Nia protein.  
 FT CHAIN 2294 2811 replicase.  
 FT CHAIN 2812 3143 coat protein.  
 SQ SEQUENCE 3143 AA; 355170 MW; 458251AA2279108E CRC64;  
 Query Match 57.3%; Score 47; DB 2; Length 3143;  
 Best Local Similarity 66.7%; Pred. No. 2.7e+02;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 PGKKVWLGSTSS 13  
 DB 311 PGKKFWVGFTNS 322  
 RESULT 20  
 Q6Y3X4\_9POTV PRELIMINARY; PRT; 3143 AA.  
 ID Q6Y3X4\_9POTV

AC O6Y3X4;  
 DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.  
 DT 05-JUL-2004, sequence version 1.  
 DT 07-FEB-2006, entry version 9.  
 DE Polyprotein.  
 OS Plum pox virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;  
 OC Potyvirus.  
 OC NCBI\_TaxID=12211;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22939292; PubMed=14579174; DOI=10.1007/s00705-003-0175-9;  
 RA Panigluolo A., Comes S., Maiss E., Piazzolla P., Crescenzi A.;  
 RT "The complete nucleotide sequence of Plum pox virus isolates from  
 RT sweet (PPV-SwC) and sour (PPV-Soc) cherry and their taxonomic  
 RT relationships within the species."; Arch. Virol. 148:2137-2153(2003).  
 RL Arch. Virol. 148:2137-2153(2003).  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 CC EMBL; AY184478; AAC62574.1; -; mRNA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.  
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR011545; DEAD/DEAH\_N.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR002540; Pept\_S30\_Pot\_P1.  
 DR InterPro; IPR001730; Peptidase\_C4.  
 DR InterPro; IPR001456; Peptidase\_C6.  
 DR InterPro; IPR001592; Poty\_coat.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR001205; RNA\_pol\_P3D.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF00270; DEAD; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00863; Peptidase\_C4; 1.  
 DR Pfam; PF00851; Peptidase\_C6; 1.  
 DR Pfam; PF01577; Peptidase\_S30; 1.  
 DR Pfam; PF00767; Poty\_coat; 1.  
 DR Pfam; PF00680; RdRP\_1; 1.  
 DR PRINTS; PR00966; NIAPOTYPTASE.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELICC; 1.  
 KW Polyprotein.  
 SQ SEQUENCE 3143 AA; 355410 MW; BEF5055767DBAE88 CRC64;  
 Query Match 57.3%; Score 47; DB 2; Length 3143;  
 Best Local Similarity 66.7%; Pred. No. 2.7e+02;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 PGKKVWLGSTSS 13  
 DB 311 PGKKFWVGFTNS 322  
 RESULT 21  
 Q858F7\_9CAUD  
 ID Q858F7\_9CAUD PRELIMINARY; PRT; 64 AA.  
 AC Q858F7;  
 DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.  
 DT 01-JUN-2003, sequence version 1.  
 DT 07-FEB-2006, entry version 6.  
 DE 18.  
 OS Enterobacteria phage epsilon15.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;

```

OC P22-like viruses.
OX NCBI_TaxID=215158;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Kropinski A.M., Billington S.J., Patrick A.N., Burts B.D.,
RA Kovalyova I., McConnell M.R.; to the EMBL/GenBank/DBJ databases.
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC
DR ENBL; AY150271; AA006081.1; -; Genomic DNA.
SQ SEQUENCE 64 AA; 7010 MW; CE8DDBB6DE85B88F CRC64;

Query Match 56.1%; Score 46; DB 2; Length 64;
Best Local Similarity 90.0%; Pred. No. 6.5;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 KKWVLGETSS 13
Db 34 KKGWLGETSS 43

RESULT 22
Q6J9N4 ARATH PRELIMINARY; PRT; 157 AA.
AC O6J9N4;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Putative AP2/EREBP transcription factor.
GN ORFNames=At2g40350;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsais.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Pan Y., Gong W., Liu D., Fu Q., Mei W.-Q., Song W.-Q., Ma L.-G.,
RA Luo J.-C., Deng X.-W., Zhu Y.-X.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC
DR EMBL; AY560891; AAT44958.1; -; mRNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001471; TF_ERF.
DR Pfam; PF00847; AP2; 1.
DR PRINTS; PD00367; ETHRSPLEMT.
DR ProDom; PD001423; TF_ERF; 1.
DR SMART; SM00380; AP2; 1.
DR PROSITE; PS1032; AP2_ERF; 1.
SQ SEQUENCE 157 AA; 17758 MW; 9B5DF7BD39B42605 CRC64;

Query Match 56.1%; Score 46; DB 2; Length 157;
Best Local Similarity 61.5%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 GKKWLGETSSAY 15
Db 88 GAKLWLGTSSSY 100

RESULT 23
Q47TS8 THEFY PRELIMINARY; PRT; 203 AA.
AC Q47TS8;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2005, sequence version 1.

```

```

DT 07-FEB-2006, entry version 4.
DE Hypothetical protein.
GN OrderedLocustNames=Tfu 0098;
OS Thermobifida fusca (strain YX).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=269800;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RG US DOE Joint Genome Institute;
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hannon N., Ibrani S., Pittluck S., Di Barcolo G., Chain P., Schmutz J.,
RA Larimer F., Land M., Lykidis A., Richardson P.;
RL "Complete sequence of Thermobifida fusca YX.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC
DR EMBL; CP000088; AA254136.1; -; Genomic DNA.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 203 AA; 21988 MW; E49349E55141ADD3 CRC64;

Query Match 56.1%; Score 46; DB 2; Length 203;
Best Local Similarity 72.7%; Pred. No. 22;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 GKKWLGETSS 13
Db 163 GYKWLGDITS 173

RESULT 24
Q7S5G5 NEUCR PRELIMINARY; PRT; 827 AA.
AC Q7S5G5;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 15-DEC-2003, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Hypothetical protein.
GN ORFNames=NCU06123.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=74-OR23-1A / FGSC 987;
RX MEDLINE=22598136; PubMed=12712197; DOI=10.1038/nature01554;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Bell-Pedersen D., Neilson M.A.,
RA Werner-Washburne M., Selitrennikoff C.P., Kinsey J.A., Braun E.L.,
RA Zelter A., Schulte U., Koth G.O., Jedd G., Mewes H.-W., Staben C.,
RA Marcotte E., Greenberg D., Roy A., Foley K., Naylor J.,
RA Stange-Thomann N., Barrett R., Gnerre S., Kamal M., Kamvysselis M.,
RA Maucelli E., Bielke C., Rudd S., Frisman D., Krystofova S.,
RA Ramussen C., Metzberg R.L., Perkins D.D., Kroken S., Cogoni C.,
RA DeSouza C.P.C., Glass N.L., Orbach M.J., Berglund J.A., Voelker R.,
RA Varden O., Plamann M., Seiler S., Dunlap J.C., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.W.;
RL "The genome sequence of the filamentous fungus Neurospora crassa.";
RL Nature 422:859-868(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----

```

```

DR EMBL; AABX01000354; EAA30749.1; -; Genomic DNA.
DR GO; GO:0016832; F:aldehyde-lyase activity; IEA.
DR CO; CO:0016829; F:lyase activity; IEA.
DR GO; GO:0030976; F:thiamin pyrophosphate binding; IEA.
DR InterPro; IPR012109; Phosphoketolase.
DR Pfam; PF03894; XFP; 1.
DR PRSF; PRSF017245; Phosphoketolase; 1.
DR PROSITE; PS60002; PHOSPHOKETOALASE_1; UNKNOWN_1.
DR PROSITE; PS60003; PHOSPHOKETOALASE_2; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 827 AA; A6D1CD869FB9B3E7 CRC64;

Query Match 56.1%; Score 46; DB 2; Length 827;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GKKVWLGE 10
    |||||
DB 453 GKKVWLGE 460

RESULT 25
EX5B_ECOLI
ID EX5B_ECOLI STANDARD; PRT; 1180 AA.
AC P08394;
DT 01-AUG-1988, integrated into UniProtKB/Swiss-Prot.
DT 01-AUG-1988, sequence version 1.
DT 07-MAR-2006, entry version 54.
DE Exodeoxyribonuclease V beta chain (EC 3.1.11.5) (Exodeoxyribonuclease
DE V 135 kDa polypeptide).
GN Name=recB; Synonyms=roxA; OrderedLocusNames=b2820;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=V1000;
RX MEDLINE=87066729; PubMed=3537960;
RA Finch P.W., Storey A., Chapman K.E., Brown K., Hickson I.D.,
RA Emerson P.T.;
RT "Complete nucleotide sequence of the Escherichia coli recB gene.";
RL Nucleic Acids Res. 14:8573-8582(1986).
RN [2]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=V1000;
RX MEDLINE=20229837; PubMed=10766864; DOI=10.1074/jbc.275.16.12261;
RA Arnold D.A., Kowalczykowski S.C.;
RT "Facilitated loading of RecA protein is essential to recombination by
RT RecBCD enzyme.";
RL J. Biol. Chem. 275:12261-12265(2000).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA] OF 1-11.
RX MEDLINE=87040734; PubMed=3534791;
RA Finch P.W., Wilson R.E., Brown K., Hickson I.D., Emerson P.T.;
RT "Complete nucleotide sequence of the Escherichia coli ptr gene
RT encoding protease III.";
RL Nucleic Acids Res. 14:7695-7703(1986).
RN [5]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA] OF 1093-1180.
RX MEDLINE=87066730; PubMed=3537961;
RA Finch P.W., Storey A., Brown K., Hickson I.D., Emerson P.T.;
RT "Complete nucleotide sequence of recD, the structural gene for the

```

```

RT alpha subunit of Exonuclease V of Escherichia coli.";
RL Nucleic Acids Res. 14:8583-8594(1986).
CC -!- FUNCTION: Required for efficient DNA repair; it catalyzes the
CC unwinding of double-stranded DNA and the cleavage of single-
CC stranded DNA and it stimulates local genetic recombination. All of
CC these activities require concomitant hydrolysis of ATP.
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of
CC ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-
CC phosphooligonucleotides.
CC -!- SUBUNIT: Consist of three subunits; recB, recC and recD.
CC -!- SIMILARITY: Belongs to the helicase family. UvrD subfamily.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
EMBL; X04581; CA28250.1; -; Genomic DNA.
EMBL; AF179304; AAD56369.1; -; Genomic DNA.
EMBL; U29581; AAB40467.1; -; Genomic DNA.
EMBL; U00096; AAC75859.1; -; Genomic DNA.
EMBL; X06227; CA229577.1; -; Genomic DNA.
EMBL; X04582; CA28252.1; -; Genomic DNA.
PIR; A25532; NCECK5.
PDB; 1W36; X-ray; B/E=1-1180.
GenomeReviews; U00096 GR; b2820.
EcoBASE; EB0817; -; EcoGene; EGI0824; recB.
BioCyc; EcoCyc:EG10824-MONOMER; -.
GO; GO:0005515; F:protein binding; IPI.
InterPro; IPR004586; RecB.
InterPro; IPR000212; UvrD-helicase.
PANTHER; PTHR11070; UvrD-helicase; 2.
Pfam; PF00580; UvrD-helicase; 1.
TIGRFAMs; TIGR00609; recB; 1.
KW 3D-structure; ATP-binding; Complete proteome; DNA damage; DNA repair;
KW Endonuclease; Exonuclease; Helicase; Hydrolase; Nuclease;
KW Nucleotide-binding.
FT CHAIN 1 1180 Exodeoxyribonuclease V beta chain.
FT NP_BIND 23 30 /FTID=PRO_0000102046.
FT STRAND 3 4
FT STRAND 6 7
FT HELIX 10 12
FT STRAND 16 17
FT STRAND 19 22
FT TURN 25 26
FT STRAND 27 27
FT HELIX 29 41
FT TURN 42 43
FT STRAND 45 49
FT STRAND 51 52
FT HELIX 56 58
FT STRAND 59 64
FT HELIX 66 88
FT STRAND 89 90
FT STRAND 94 94
FT STRAND 95 103
FT STRAND 105 105
FT HELIX 107 123
FT STRAND 125 128
FT STRAND 129 139
FT TURN 140 140
FT HELIX 141 144
FT TURN 145 145
FT STRAND 148 149
FT STRAND 151 152
FT HELIX 157 172
FT TURN 173 174
FT HELIX 177 186
FT STRAND 188 189
FT STRAND 190 197
FT TURN 198 201
FT STRAND 202 204
FT STRAND 207 210

```

FT TURN 214 215  
FT TURN 218 234  
FT TURN 235 236  
FT STRAND 237 237  
FT STRAND 240 241  
FT STRAND 249 251  
FT STRAND 253 253  
FT TURN 254 254  
FT HELIX 255 263  
FT TURN 264 264  
FT STRAND 267 268  
FT STRAND 272 272  
FT TURN 279 280  
FT STRAND 281 282  
FT HELIX 283 287  
FT STRAND 309 309  
FT TURN 311 311  
FT HELIX 312 316  
FT TURN 317 318  
FT STRAND 319 319  
FT STRAND 323 323  
FT STRAND 324 346  
FT TURN 347 347  
FT STRAND 348 349  
FT HELIX 351 363  
FT STRAND 364 364  
FT TURN 365 366  
FT HELIX 367 377  
FT STRAND 379 383  
FT STRAND 385 385  
FT HELIX 386 388  
FT HELIX 391 401  
FT TURN 402 403  
FT TURN 405 406  
FT STRAND 408 413  
FT HELIX 415 417  
FT TURN 421 424  
FT STRAND 427 436  
FT STRAND 440 442  
FT STRAND 445 446  
FT TURN 447 448  
FT STRAND 449 449  
FT HELIX 451 462  
FT STRAND 463 466  
FT TURN 467 468  
FT STRAND 469 470  
FT TURN 471 472  
FT STRAND 473 473  
FT STRAND 479 480  
FT HELIX 482 484  
FT TURN 485 486  
FT STRAND 487 491  
FT TURN 492 493  
FT STRAND 494 496  
FT STRAND 498 503  
FT STRAND 506 507  
FT TURN 511 512  
FT HELIX 513 534  
FT TURN 535 536  
FT STRAND 538 542  
FT TURN 543 544  
FT STRAND 545 548  
FT HELIX 551 553  
FT STRAND 554 560  
FT HELIX 561 572  
FT TURN 573 575  
FT STRAND 578 580  
FT TURN 581 582  
FT STRAND 586 586  
FT HELIX 587 589  
FT STRAND 590 591  
FT HELIX 592 603  
FT TURN 604 604

FT TURN 606 607  
FT STRAND 608 608  
FT HELIX 609 617  
FT STRAND 618 618  
FT HELIX 619 621

Query Match 56.1%; Score 46; DB 1; Length 1180;  
Best Local Similarity 75.0%; Pred. NO. 1.4e+02;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 KKWLGESSAY 15  
| | | | |  
Db 1082 KSNWLGESSAY 1093

RESULT 26  
Q31XG8\_SHIDS  
ID Q31XG8\_SHIDS PRELIMINARY; PRT; 1180 AA.  
AC Q31XG8;  
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.  
DT 06-DEC-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE ATP-dependent dsDNA/ssDNA exonuclease V subunit.  
GN Name=recB; OrderedLocustNames=SBO\_2710;  
OS Shigella boydii serotype 4 (strain Sb227).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Shigella.  
OX NCBI\_TaxID=300268;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RX PubMed=16275786; DOI=10.1093/nar/gki954;  
RA Yang F., Yang J., Zhang X., Chen L., Jiang Y., Yan Y., Tang X.,  
Wang J., Xiong Z., Dong J., Xue Y., Xu X., Sun L., Chen S.,  
Nie H., Peng J., Xu J., Wang Y., Yuan Z., Wen Y., Yao Z., Shen Y.,  
Qiang B., Hou Y., Yu J., Jin Q.;  
RA "Genome dynamics and diversity of Shigella species, the etiologic  
agents of bacillary dysentery.";  
RL Nucleic Acids Res. 33:6445-6458(2005).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; CP000036; AB67240.1; -; Genomic\_DNA.  
DR GO; GO:000524; F:ATP binding; IEA.  
DR GO; GO:0004003; F:ATP-dependent DNA helicase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008854; F:exodeoxyribonuclease V activity; IEA.  
DR GO; GO:0004527; F:exonuclease activity; IEA.  
DR GO; GO:0006281; P:DNA repair; IEA.  
KW Complete proteome; Exonuclease.  
SQ SEQUENCE 1180 AA; 133973 MW; 184724E111BDDAAB CRC64;

Query Match 56.1%; Score 46; DB 2; Length 1180;  
Best Local Similarity 75.0%; Pred. NO. 1.4e+02;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 KKWLGESSAY 15  
| | | | |  
Db 1082 KSNWLGESSAY 1093

RESULT 27  
Q32CA1\_SHIDS  
ID Q32CA1\_SHIDS PRELIMINARY; PRT; 1180 AA.  
AC Q32CA1;  
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.  
DT 06-DEC-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE ATP-dependent dsDNA/ssDNA exonuclease V subunit.  
GN Name=recB; OrderedLocustNames=SDY\_3037;  
OS Shigella dysenteriae serotype 1 (strain Sd197).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Shigella.



OX NCBI\_TaxID=300267;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RX PubMed=16275786; DOI=10.1093/nar/gki954;  
RA Yang F., Yang J., Zhang X., Chen L., Jiang Y., Yan Y., Tang X.,  
RA Wang J., Xiong Z., Dong J., Xue Y., Zhu Y., Xu X., Chen S.,  
RA Nie H., Feng J., Xu J., Wang Y., Yuan Z., Wen Y., Yao Z., Shen Y.,  
RA Qiang B., Hou Y., Yu J., Jin Q.;  
RT "Genome dynamics and diversity of Shigella species, the etiologic  
RT agents of bacillary dysentery.";  
RL Nucleic Acids Res. 33:6445-6458(2005).  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL: CP000034; AB63054.1; -; Genomic DNA.  
DR GO: 0005524; F:ATP binding; IEA.  
DR GO: 0004003; F:ATP-dependent DNA helicase activity; IEA.  
DR GO: 0003677; F:DNA binding; IEA.  
DR GO: 0008854; F:exodeoxyribonuclease V activity; IEA.  
DR GO: 0004527; F:exonuclease activity; IEA.  
DR GO: 0006281; F:DNA repair; IEA.  
KW Complete proteome; Exonuclease.  
SQ SEQUENCE 1180 AA; 133994 MW; 119D5FF6F2D04EC2 CRC64;  
  
Query Match 56.1%; Score 46; DB 2; Length 1180;  
Best Local Similarity 75.0%; Pred. No. 1.4e+02;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 4 KKVWLGTSAY 15  
Db 1082 KSNWLGDSAY 1093  
  
RESULT 28  
Q3YY40 SHISS  
ID Q3YY40\_SHISS PRELIMINARY; PRT; 1180 AA.  
AC Q3YY40;  
DT 27-SEP-2005, integrated into UniProtKB/TrEMBL.  
DT 27-SEP-2005, sequence version 1.  
DT 07-FEB-2006, entry version 5.  
DE DNA helicase ATP-dependent dcdna/ssDNA exonuclease V subunit.  
GN Name=recB; OrderedLocNames=SSO\_2977; ORFNames=SSO\_2977;  
OS Shigella sonnei (strain S9046).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Shigella.  
OX NCBI\_TaxID=300269;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RX PubMed=16275786; DOI=10.1093/nar/gki954;  
RA Yang F., Yang J., Zhang X., Chen L., Jiang Y., Yan Y., Tang X.,  
RA Wang J., Xiong Z., Dong J., Xue Y., Zhu Y., Xu X., Chen S.,  
RA Nie H., Feng J., Xu J., Wang Y., Yuan Z., Wen Y., Yao Z., Shen Y.,  
RA Qiang B., Hou Y., Yu J., Jin Q.;  
RT "Genome dynamics and diversity of Shigella species, the etiologic  
RT agents of bacillary dysentery.";  
RL Nucleic Acids Res. 33:6445-6458(2005).  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL: CP000038; AA289572.1; -; Genomic DNA.  
DR GO: 0005524; F:ATP binding; IEA.  
DR GO: 0004003; F:ATP-dependent DNA helicase activity; IEA.  
DR GO: 0003677; F:DNA binding; IEA.  
DR GO: 0008854; F:exodeoxyribonuclease V activity; IEA.  
DR GO: 0004527; F:exonuclease activity; IEA.  
DR GO: 0004386; F:helicase activity; IEA.  
DR GO: 0006281; P:DNA repair; IEA.  
DR InterPro: IPR004586; RecB.  
DR InterPro: IPR000212; UvrD-helicase.  
DR Pfam: PF00580; UvrD-helicase; 1.  
DR TIGRFAMs: TIGR00609; recB; 1.

KW Complete proteome; Exonuclease; Helicase.  
SQ SEQUENCE 1180 AA; 133972 MW; 28B03B16E6341FEB CRC64;  
  
Query Match 56.1%; Score 46; DB 2; Length 1180;  
Best Local Similarity 75.0%; Pred. No. 1.4e+02;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 4 KKVWLGTSAY 15  
Db 1082 KSNWLGDSAY 1093  
  
RESULT 29  
Q2MA17 ECOLI  
ID Q2MA17\_ECOLI PRELIMINARY; PRT; 1180 AA.  
AC Q2MA17;  
DT 21-FEB-2006, integrated into UniProtKB/TrEMBL.  
DT 21-FEB-2006, sequence version 1.  
DT 07-MAR-2006, entry version 2.  
DE Exonuclease V (RecBCD complex), beta subunit.  
GN Name=recB;  
OS Escherichia coli W3110.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=316407;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=K-12;  
RX MEDLINE=81053692; PubMed=6159575;  
RA Smith D.R., Calvo J.M.;  
RT "Nucleotide sequence of the E coli gene coding for dihydrofolate  
RT reductase.";  
RL Nucleic Acids Res. 8:2255-2274 (1980).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=K-12;  
RA Sekiguchi T., Ortega-Cesena J., Nosoh Y., Ohashi S., Tsuda K.,  
RA Kanaya S.;  
RT "DNA and amino-acid sequences of 3-isopropylmalate dehydrogenase of  
RT Bacillus coagulans. Comparison with the enzymes of Saccharomyces  
RT cerevisiae and Thermus thermophilus.";  
RL Biochim. Biophys. Acta 867:36-44(1986).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=K-12;  
RA Chen H., Sun Y., Stark T., Beattie W., Moses R.E.;  
RT "Nucleotide sequence and deletion analysis of the polB gene of  
RT Escherichia coli.";  
RL DNA Cell Biol. 9:613-635(1990).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=K-12;  
RA Smallshaw J.E., Kellin R.A.;  
RT "Cloning, nucleotide sequence and expression of the Escherichia coli  
RT K-12 pyrH gene encoding UMP kinase.";  
RL Genetics (Life Sci. Adv.) 11:59-65(1992).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=K-12;  
RA Hayashi K., Morooka N., Yamamoto Y., Fujita K., Isono K., Choi S.,  
RA Ohtsubo E., Baba T., Wanner B.L., Mori H., Horiuchi T.;  
RT "Highly accurate genome sequences of Escherichia coli K-12 strains  
RT MG1655 and W3110.";  
RL Mol. Syst. Biol. 0:0-0(2006).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=K-12;  
RX PubMed=16397293; DOI=10.1093/nar/gkj150;  
RA Riley M., Abe T., Arnaud M.B., Berlyn M.K., Blattner F.R.,  
RA Chaudhuri R.R., Glasner J.D., Horiuchi T., Keseler I.M., Kosuge T.,  
RA Mori H., Perna N.T., Plunkett G. III, Rudd K.E., Serres M.H.,  
RA Thomas G.H., Thomson N.R., Wishart D., Wanner B.L.;  
RT "Escherichia coli K-12: a cooperatively developed annotation snapshot-

RT -2005.";  
RL Nucleic Acids Res. 34:1-9(2006).  
RN [7]  
RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=K-12;  
RX MEDLINE=97349980; PubMed=9205837; DOI=10.1093/dnares/4.2.91;  
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,  
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto S.,  
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,  
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
RA Yamagata S., Horiuchi T.;  
RT "Construction of a contiguous 874-kb sequence of the *Escherichia coli*  
RT -K12 genome corresponding to 50.0-68.8 min on the linkage map and  
RT analysis of its sequence features";  
RL DNA Res. 4:91-113(1997).  
RN [8]  
RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=K-12;  
RX MEDLINE=97251358; PubMed=9097040; DOI=10.1093/dnares/3.6.379;  
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,  
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,  
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,  
RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,  
RA Sivasundaram S., Tagami H., Takeda J., Takemoto K., Wada C.,  
RA Yamamoto Y., Horiuchi T.;  
RT "A 460-kb DNA sequence of the *Escherichia coli* K-12 genome  
RT corresponding to the 40.1-50.0 min region on the linkage map";  
RL DNA Res. 3:379-392(1996).  
RN [9]  
RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=K-12;  
RX MEDLINE=97251357; PubMed=9097039; DOI=10.1093/dnares/3.6.363;  
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,  
RA Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,  
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;  
RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome  
RT corresponding to the 28.0-40.1 min region on the linkage map";  
RL DNA Res. 3:363-377(1996).  
RN [10]  
RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=K-12;  
RX MEDLINE=97094878; PubMed=8940112; DOI=10.1074/jbc.271.49.31145;  
RA Arn E.A., Abelson J.N.;  
RT "The 2'-5' RNA ligase of *Escherichia coli*. Purification, cloning, and  
RT genomic disruption";  
RL J. Biol. Chem. 271:31145-31153(1996).  
RN [11]  
RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=K-12;  
RX MEDLINE=97061202; PubMed=8905232; DOI=10.1093/dnares/3.3.137;  
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,  
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome  
RT corresponding to the 12.7-28.0 min region on the linkage map";  
RL DNA Res. 3:137-155(1996).  
RN [12]  
RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=K-12;  
RX MEDLINE=94561430; PubMed=8202364;  
RA Fujita N., Mori H., Yura T., Ishihara A.;  
RT "Systematic sequencing of the *Escherichia coli* genome: analysis of the  
RT 2.4-4.1 min (110,917-193,643 bp) region";  
RL Nucleic Acids Res. 22:1637-1639(1994).  
RN [13]  
RC NUCLEOTIDE SEQUENCE.

RC STRAIN=K-12;  
RX MEDLINE=94240115; PubMed=8183897;  
RA Janosi L., Shimizu I., Kaji A.;  
RT "Ribosome recycling factor (ribosome releasing factor) is essential  
RT for bacterial growth";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:4249-4253(1994).  
RN [14]  
RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=K-12;  
RX MEDLINE=94124004; PubMed=7904973; DOI=10.1016/0378-1119(93)90470-N;  
RA Allknetts R., Gerrard B.C., Court D., Dean M.C.;  
RT "Cloning and organization of the *abc* and *mdl* genes of *Escherichia*  
RT *coli*: relationship to eukaryotic multidrug resistance";  
RL Gene 136:231-236(1993).  
RN [15]  
RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=K-12;  
RX MEDLINE=94018640; PubMed=8412694;  
RA van Heeswijk W.C., Rabenberg M., Westerhoff H.V., Kahn D.D.;  
RT "The genes of the glutamine synthetase adenylylation cascade are not  
RT regulated by nitrogen in *Escherichia coli*";  
RL Mol. Microbiol. 9:443-458(1993).  
RN [16]  
RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=K-12;  
RX MEDLINE=93259920; PubMed=8387990;  
RA Zhao S., Sandt C.H., Feulner G., Vlazny D.A., Gray J.A., Hill C.W.;  
RT "Rhs elements of *Escherichia coli* K-12: complex composites of shared  
RT and unique components that have different evolutionary histories";  
RL J. Bacteriol. 175:2799-2808(1993).  
RN [17]  
RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=K-12;  
RX MEDLINE=93123180; PubMed=8419307;  
RA Yamada M., Azaoka S., Saier M.H. Jr., Yamada Y.;  
RT "Characterization of the *gcd* gene from *Escherichia coli* K-12 W3110 and  
RT regulation of its expression";  
RL J. Bacteriol. 175:568-571(1993).  
RN [18]  
RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=K-12;  
RX MEDLINE=93116053; PubMed=1474579;  
RA Cormack R.S., Mackie G.A.;  
RT "Structural requirements for the processing of *Escherichia coli* 5 S  
RT ribosomal RNA by RNase E in vitro";  
RL J. Mol. Biol. 228:1078-1090(1992).  
RN [19]  
RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=K-12;  
RX MEDLINE=93094132; PubMed=1459951;  
RA Gervais F.G., Drapeau G.R.;  
RT "Identification, cloning, and characterization of *rscF*, a new  
RT regulator gene for exopolysaccharide synthesis that suppresses the  
RT division mutation *fts284* in *Escherichia coli* K-12";  
RL J. Bacteriol. 174:8016-8022(1992).  
RN [20]  
RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=K-12;  
RX MEDLINE=93077430; PubMed=1447125;  
RA Yananaka K., Ogura T., Niki H., Hiraga S.;  
RT "Identification and characterization of the *smbA* gene, a suppressor of  
RT the *mukB* null mutant of *Escherichia coli*";  
RL J. Bacteriol. 174:7517-7526(1992).  
RN [21]  
RC NUCLEOTIDE SEQUENCE.  
RP STRAIN=K-12;  
RX MEDLINE=93011013; PubMed=1396599;  
RA Condon C., Phillips J., Fu Z.Y., Squires C., Squires C.L.;  
Query Match 56.1%; Score 46; DB 2; Length 1180;  
Best Local Similarity 75.0%; Pred. NO. 1.4e+02;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 KKVWLGETSSAY 15  
| | | | |  
DB 1082 KSNWLGEDSSAY 1093

```
RESULT 30
Q8XGM9 ECOS7
ID Q8XGM9 ECOS7 PRELIMINARY; PRT; 1180 AA.
AC Q8XGM9; 07AB60;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2002, sequence version 1.
DT 07-FEB-2006, entry version 24.
DE DNA helicase, ATP-dependent dsDNA/ssDNA exonuclease V subunit, ssDNA
DE endonuclease (DNA helicase RecB).
GN Name=recB; OrderedLocusNames=ECs3677, z1137;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RX MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glaesner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grofbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=O157:H7 / Sakai / RIMD 050952 / EHEC;
RX MEDLINE=21156231; PubMed=1258796; DOI=10.1093/dnares/8.1.11;
RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; BA000007; BAB37100.1; -; Genomic DNA.
DR EMBL; AE005174; AAG57931.1; -; Genomic DNA.
DR PIR; E91088; E91088.
DR PIR; G85933; G85933.
DR SMR; Q8XGM9; 1-1174.
DR BioCyc; ECOL83334-1:ECs3677-MONOMER; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004003; F:ATP-dependent DNA helicase activity; IEA.
DR GO; GO:0004003; F:ATP-dependent DNA helicase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0008854; F:exodeoxyribonuclease V activity; IEA.
DR GO; GO:0004527; F:exonuclease activity; IEA.
DR GO; GO:0004386; F:helicase activity; IEA.
DR GO; GO:0006281; F:DNA repair; IEA.
DR InterPro; IPR004586; RecB.
DR InterPro; IPR000212; UvrD-helicase.
DR PANTHER; PTHR11070; UvrD-helicase; 2.
DR Pfam; PF00580; UvrD-helicase; 1.
DR TIGRFAMs; TIGR00609; recB; 1.
DR TIGRFAMs; TIGR00609; recB; 1.
KW Complete proteome; Endonuclease; Exonuclease; Helicase.
SQ SEQUENCE 1180 AA; 134111 MW; D8CE1569E46F5F65 CRC64;
```

Query Match 56.1%; Score 46; DB 2; Length 1180;  
Best Local Similarity 75.0%; Pred. No. 1.4e+02;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 KKVWLGETSSAY 15

DB 1082 KSNWLGEDSSAY 1093  
| | | | |

```
RESULT 31
Q83JW0 SHIFL
ID Q83JW0 SHIFL PRELIMINARY; PRT; 1180 AA.
AC Q83JW0_07UBS0;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-FEB-2005, sequence version 2.
DT 07-MAR-2006, entry version 19.
DE DNA helicase, ATP-dependent dsDNA/ssDNA exonuclease V subunit, ssDNA
DE endonuclease.
GN Name=recB; OrderedLocusNames=S3028, SF2831; ORFNames=S_3028;
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590; DOI=10.1093/nar/gkf566;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
DOI=10.1128/IAI.71.5.2775-2786.2003;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; AE005674; AAN44318.2; -; Genomic DNA.
DR EMBL; AE014073; AAP18143.1; -; Genomic DNA.
DR SMR; Q83JW0; 1-1174.
DR GenomeReviews; AE014073 GR; S3028.
DR GenomeReviews; AE005674 GR; SF2831.
DR BioCyc; SFLE198214:AA44318.1-MO-; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004003; F:ATP-dependent DNA helicase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0008854; F:exodeoxyribonuclease V activity; IEA.
DR GO; GO:0004527; F:exonuclease activity; IEA.
DR GO; GO:0004386; F:helicase activity; IEA.
DR GO; GO:0006281; F:DNA repair; IEA.
DR InterPro; IPR004586; RecB.
DR InterPro; IPR000212; UvrD-helicase.
DR PANTHER; PTHR11070; UvrD-helicase; 2.
DR Pfam; PF00580; UvrD-helicase; 1.
DR TIGRFAMs; TIGR00609; recB; 1.
KW Complete proteome; Endonuclease; Exonuclease; Helicase.
SQ SEQUENCE 1180 AA; 134044 MW; A471BC694EC0D433 CRC64;
```

Query Match 56.1%; Score 46; DB 2; Length 1180;  
Best Local Similarity 75.0%; Pred. No. 1.4e+02;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 KKVWLGETSSAY 15

```
Db      1082 KSNWLGEDSSAY 1093
|      ||||| |||||
RESULT 32
Q8FEB3 ECOL6 PRELIMINARY; PRT; 1183 AA.
AC Q8FEB3;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 21-FEB-2006, entry version 16.
DE Exodeoxyribonuclease V beta chain (EC 3.1.11.5).
GN Name=recB; ORFNames=C_3414;
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=06:H1 / CFT073 / ATCC 700928 / UPEC;
RX MEDLINE=22388234; PubMed=12471157; DOI=10.1073/pnas.252529799;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; A2014075; AAN81859.1; -; Genomic_DNA.
DR SMR; Q8FEB3; 4-1177.
DR BioCyc; ECOL199310:C3414-MONOMER; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004003; F:ATP-dependent DNA helicase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003854; F:exodeoxyribonuclease V activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR InterPro; IPR004586; RecB.
DR PANTHER; PTHR11070; UvrD-helicase.
DR Pfam; PF00580; UvrD-helicase; 1.
DR TIGRFAMs; TIGR00609; recB; 1.
KW Complete proteome; Hydrolase.
SQ SEQUENCE 1183 AA; 134424 MW; F11CB6919C16B8CA CRC64;

Query Match 56.1%; Score 46; DB 2; Length 1183;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      4 KKWLGETSSAY 15
|      ||||| |||||
Db      1085 KSNWLGEDSSAY 1096
|      ||||| |||||
RESULT 33
Q8YKV0 ANASP PRELIMINARY; PRT; 1596 AA.
AC Q8YKV0;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2002, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE All7191 protein.
GN OrderedLocNames=all7191;
OS Anabaena sp. (strain PCC 7120).
OG Plasmid pCC7120alpha.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

Query Match 56.1%; Score 46; DB 2; Length 1183;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
RX MEDLINE=21595285; PubMed=11759840; DOI=10.1093/dnares/8.5.205;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213 (2001).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BA000020; BAB78275.1; -; Genomic_DNA.
DR PIR; AG2501; AG2501.
DR BioCyc; NSP103690:ALL7191-MONOMER; -.
KW Complete proteome; Plasmid.
SQ SEQUENCE 1596 AA; 180618 MW; F812AD045822D289 CRC64;

Query Match 55.5%; Score 45; DB 2; Length 1596;
Best Local Similarity 52.9%; Pred. No. 2.4e+02;
Matches 9; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

Qy      1 RPQ---KKWLGETSSA 14
|      ||| :||| | :
Db      240 RPGLYKQIKWLGERKSQ 256
|      ||| :||| | :

RESULT 34
Q3P129 PARDE PRELIMINARY; PRT; 73 AA.
AC Q3P129;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Hypothetical protein.
GN ORFNames=PdenDRAFT_4242;
OS Paracoccus denitrificans PD1222.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Paracoccus.
OX NCBI_TaxID=318586;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PD1222;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Paracoccus
denitrificans PD1222.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PD1222;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Paracoccus denitrificans
PD1222.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC !- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AA10100002; EAM67635.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 73 AA; 8145 MW; 8C3DF87B487EAFB3 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 73;
Best Local Similarity 57.1%; Pred. No. 11;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 1 RPKKVLGETSSA 14
Db 20 RDGKVRGKCASA 33

RESULT 35
Q44177_SYPN2 PRELIMINARY; PRT; 134 AA.
AC Q44177;
DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1996, sequence version 1.
DT 07-FEB-2006, entry version 19.
DE ORF134.
OS Synechococcus PCC7002 PR-6.
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=34076;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PCC 7002;
RA Akiyama H., Kanai S., Hirano M., Sugimoto M., Kiyohara M.;
RL Submitted (DRC-1992) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL: D13971; BAA03077.1; -; Genomic_DNA.
DR InterPro; IPR003435; Chaperonin_RcbX.
DR Pfam; PF02341; RcbX; 1.
SQ SEQUENCE 134 AA; 15269 MW; A11F4296F424988 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 134;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 PGKQVLGETSSAY 15
Db 35 PGQAVLGEFSKRH 48

RESULT 36
DRE2H ARATH STANDARD; PRT; 164 AA.
ID DRE2H ARATH
AC Q8S1Z0;
DT 07-JUN-2004, integrated into UniProtKB/Swiss-Prot.
DT 07-JUN-2004, sequence version 2.
DT 07-MAR-2006, entry version 31.
DE Putative dehydration-responsive element-binding protein 2H (DREB2H
DE protein).
GN Names=DREB2H; OrderedLocustNames=At2g40350; ORFNames=T3G21.12;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197; DOI=10.1038/45471;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.-J., Rensing C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RA thaliana."
RL Nature 402:761-768(1999).
RN [2]
RP GENE FAMILY, AND FUNCTION.
RX MEDLINE=21656975; PubMed=11798174; DOI=10.1006/bbrc.2001.6299;
RA Sakuma Y., Liu Q., Dubouzet J.G., Abe H., Shinozaki K.,
```

```
RA Yamaguchi-Shinozaki K.;
RT "DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs,
RT transcription factors involved in dehydration- and cold-inducible gene
RT expression.";
RL Biochem. Biophys. Res. Commun. 290:998-1009(2002).
CC -!- FUNCTION: Putative transcriptional activator that binds
CC specifically to the DNA sequence 5'-[AG]CGAC-3' (By similarity).
CC -!- SUBCELLULAR LOCATION: Nucleus (probable).
CC -!- SIMILARITY: Contains 1 AP2/ERF DNA-binding domain.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to erroneous
CC gene model prediction. May be the product of a pseudogene.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL: AC007020; AAD25670.1; ALT_SEQ; Genomic_DNA.
DR PIR; D84828; D84828.
DR HSP; O80337; 2GCC.
DR GeneFarm; 3683; 373.
DR TAIR; At2g40350; -.
DR InterPro; IPR001471; TF_ERF.
DR Pfam; PF00847; AP2; 1.
DR PRINTS; PR00367; ETRSPLEMT.
DR ProDom; PD001423; TF_ERF; 1.
DR SMART; SM00380; AP2; 1.
DR PROSITE; PS51032; AP2_ERF; 1.
KW Activator; DNA-binding; Hypothetical protein; Nuclear protein;
KW Transcription; Transcription regulation.
FT CHAIN 1..164
FT Binding protein 2H.
FT /FTID=PRO_0000112541.
FT DNA_BIND 73..130 AP2/ERF.
FT MOTIF 10..50 Nuclear localization signal (Potential).
FT SEQUENCE 164 AA; 18493 MW; D816D272C9DFB8 CRC64;

Query Match 54.9%; Score 45; DB 1; Length 164;
Best Local Similarity 61.5%; Pred. No. 26;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 GKKVWLGETSSAY 15
Db 95 GAKVLGTFSY 107

RESULT 37
Q5BK39 RAT PRELIMINARY; PRT; 225 AA.
ID Q5BK39 RAT
AC Q5BK39;
DT 12-APR-2005, integrated into UniProtKB/TrEMBL.
DT 12-APR-2005, sequence version 1.
DT 07-FEB-2006, entry version 7.
DE Hypothetical protein LOC308990.
GN Name=LOC308990;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Thymus;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udgin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
```

RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Thymus;  
RG NIH MGC Project;  
RG Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; BC091216; AAH91216.1; -; mRNA.  
DR EMBL; BC091205; AAH91205.1; -; mRNA.  
KW Hypothetical protein.  
SQ SEQUENCE 225 AA; 24504 MW; 8158AD445C032754 CRC64;  
  
Query Match 54.9%; Score 45; DB 2; Length 225;  
Best Local Similarity 57.1%; Pred. No. 37;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 RPQKKVWLGETSSA 14  
||| :|: |||  
Db 76 RPQGLWIEPTSSA 89  
  
RESULT 38  
Q6P3A6\_MOUSE PRELIMINARY; PRT; 225 AA.  
ID Q6P3A6\_MOUSE  
AC Q6P3A6;  
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.  
DT 05-JUL-2004, sequence version 1.  
DT 07-FEB-2006, entry version 12.  
DE Hypothetical protein A1467606.  
GN Name=A1467606;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;

RG NIH MGC Project;  
RG Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; BC064101; AAH64101.1; -; mRNA.  
DR EMBL; ENSMUSG0000045165; Mus musculus.  
KW Hypothetical protein.  
SQ SEQUENCE 225 AA; 24489 MW; 8EB83863E92FC38 CRC64;  
  
Query Match 54.9%; Score 45; DB 2; Length 225;  
Best Local Similarity 57.1%; Pred. No. 37;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
QY 1 RPQKKVWLGETSSA 14  
||| :|: |||  
Db 76 RPQGLWIEPTSSA 89  
  
RESULT 39  
Q8C708\_MOUSE PRELIMINARY; PRT; 225 AA.  
ID Q8C708\_MOUSE  
AC Q8C708;  
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
DT 01-MAR-2003, sequence version 1.  
DT 07-FEB-2006, entry version 19.  
DE 0 day neonate kidney cDNA, RIKEN full-length enriched library,  
DE clone: D630041B17 product: hypothetical protein, full insert sequence  
DE (NOD-derived CD11c +ve dendritic cells cDNA, RIKEN full-length  
DE enriched library, clone: F630208N22 product: hypothetical protein, full  
DE insert sequence) (Expressed sequence A1467606) (CRL-1722 L5178Y-R  
DE cDNA, RIKEN full-length enriched library, clone: I730073I12  
DE product: hypothetical protein, full insert sequence) (2 days neonate  
DE thymus thymic cells cDNA, RIKEN full-length enriched library,  
DE clone: E430005013 product: hypothetical protein, full insert sequence)  
DE (NOD-derived CD11c +ve dendritic cells cDNA, RIKEN full-length  
DE enriched library, clone: F630035E07 product: hypothetical protein, full  
DE insert sequence).  
GN Name=A1467606;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J, NOD, and DBA/2; TISSUE=Kidney, and Thymus;  
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.,  
RT "High-efficiency full-length cDNA cloning.";  
RN Methods Enzymol. 303:19-44 (1999).  
[2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J, NOD, and DBA/2; TISSUE=Kidney, and Thymus;  
RX PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
RA di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Heminger P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Huminicki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Keiso J., Kitamura H.,  
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,



RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlinghaus V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J.C., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Sample C.A., Seno S., Sesaa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugiyura K., Sultana R., Takenaka Y., Taki K.,  
RA Tamajima K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Wahlestedt C., Mattick J.S., Hume D.A., Kal C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563(2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J, NOD, and DBA/2; TISSUE=Kidney, and Thymus;  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
RG (Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome.";  
RL Science 309:1564-1566(2005).  
RN [4]  
RN NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney, and Thymus; STRAIN=C57BL/6J, NOD, and DBA/2;  
RX MEDLINE=22354683; PubMed=14266851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikazaki I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Mateuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawagawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reid D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Sample C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tonita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Sasaki D., Shibata K., Shinagawa A.,  
RA Birney E., Hayashizaki Y.,  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
RN [5]  
RN NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J, NOD, and DBA/2; TISSUE=Kidney, and Thymus;  
RX MEDLINE=22085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fushimi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Stabli P., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Maehina J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [6]  
RN NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J, NOD, and DBA/2; TISSUE=Kidney, and Thymus;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugihara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [7]  
RN NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J, NOD, and DBA/2; TISSUE=Kidney, and Thymus;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Hatada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwara S., Inoue K., Ozawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-Format  
RT sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [8]  
RN NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Kidney;  
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
RA Saito R., Saitoh H., Sakai K., Sakai K., Sakazume N., Sano H.,  
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
RN [9]  
RN NUCLEOTIDE SEQUENCE.  
RC STRAIN=NOD;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
RN [10]  
RN NUCLEOTIDE SEQUENCE.  
RC TISSUE=Mammary Gland;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,



```
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

Query Match          54.9%; Score 45; DB 2; Length 225;
Best Local Similarity 57.1%; Pred. No. 37; Mismatches 0; Gaps 0;
Matches 8; Conservative 3; Indels 3;

Qy 1 RPKGKVKWLGETSSA 14
    ||| : : | : |||
Db 76 RFGELWIEPTSSA 89

RESULT 40
Q8A4U3 BACTN          PRELIMINARY; PRT; 226 AA.
AC Q8A4U3;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Hypothetical protein.
GN OrderedLocusNames=BT2504; ORFNames=BT_2504;
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroidetes (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
[1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12663928; DOI=10.1126/science.1080029;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
RL Science 298:2074-2076(2003).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AS015928; AA077611.1; -; Genomic_DNA.
DR Biocyc; BTHE26186:BT2504-MONOMER; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 226 AA; 26012 MW; 98FF9C9827FE6A71 CRC64;

Query Match          54.9%; Score 45; DB 2; Length 226;
Best Local Similarity 70.8%; Pred. No. 37; Mismatches 0; Gaps 0;
Matches 7; Conservative 2; Indels 1;

Qy 1 RPKGKVKWLGE 10
    ||| : : | : |||
Db 71 RFGERVWKGE 80

RESULT 41
Q3ITU3 NATPD          PRELIMINARY; PRT; 331 AA.
AC Q3ITU3;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 1.
DT 07-MAR-2006, entry version 6.
DE Hypothetical protein.
GN OrderedLocusNames=NP0698A;
OS Natronomonas pharaonis (Strain DSM 2160 / ATCC 35678).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Natronomonas.
OX NCBI_TaxID=348780;
[1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=16169924; DOI=10.1101/gr.3952905;
RA Falb M., Pfeiffer F., Palm P., Rodewald K., Hickmann V., Tittor J.,
RA Oesterhelt D.;
RT "Living with two extremes: conclusions from the genome sequence of
RT Natronomonas pharaonis.";
```

```
RL Genome Res. 15:1336-1343(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CR936257; CAI48440.1; -; Genomic_DNA.
DR GenomeReviews; CR936257.GR; NP0698A.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 331 AA; 37126 MW; EC6E91B8FF0462AE CRC64;

Query Match          54.9%; Score 45; DB 2; Length 331;
Best Local Similarity 53.8%; Pred. No. 55; Mismatches 0; Gaps 0;
Matches 7; Conservative 2; Indels 4;

Qy 1 RPKGKVKWLGETSS 13
    ||| : : | : |||
Db 65 RPRAYWLGNTET 77

RESULT 42
Q6FA89 ACIAD          PRELIMINARY; PRT; 335 AA.
AC Q6FA89;
DT 19-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Hypothetical protein.
GN OrderedLocusNames=ACIAD2229;
OS Acinetobacter sp. (strain ADP1).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Acinetobacter.
OX NCBI_TaxID=62977;
[1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15514110; DOI=10.1093/nar/gkh910;
RA Barbe V., Vallenet D., Fonknechten N., Kreimeyer A., Oztas S.,
RA Labarre L., Cruveiller S., Robert C., Duprat S., Wincker P.,
RA Ornstom L.N., Weissbach J., Marliere P., Cohen G.N., Medigue C.;
RT "Unique features revealed by the genome sequence of Acinetobacter sp.
RT ADP1, a versatile and naturally transformation competent bacterium.";
RL Nucleic Acids Res. 32:5766-5779(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CR543861; CAG69024.1; -; Genomic_DNA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR000312; Glyco trans.3.
DR Pfam; PF02885; Glycos trans 3N; 1.
DR Pfam; PF00591; Glycos_transf.3; 1.
DR ProDom; PD001864; Glyco_trans_3; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 335 AA; 38202 MW; EF183231512EDBAB CRC64;

Query Match          54.9%; Score 45; DB 2; Length 335;
Best Local Similarity 66.7%; Pred. No. 56; Mismatches 0; Gaps 0;
Matches 8; Conservative 1; Mismatches 3; Indels 3;

Qy 4 KKVWLGETSSAY 15
    ||| : : | : |||
Db 283 KSVWLGESSHVEY 294

RESULT 43
Q5XK86 XENLA          PRELIMINARY; PRT; 473 AA.
AC Q5XK86;
DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.
DT 23-NOV-2004, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).;
```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae; Xenopus; Xenopus.  
 OC NCBI\_TaxID=8155;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Embryo;  
 RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;  
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
 RA Richardson P.;  
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
 RT initiative";  
 RL Dev. Dyn. 225:384-391(2002).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Embryo;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Embryo;  
 RA Klein S., Gerhard D.S.;  
 RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL: BC083026; AAH83026.1; -; mRNA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 473 AA; 17EE6661F76CE267 CRC64;  
 Query Match 54.9%; Score 45; DB 2; Length 473;  
 Best Local Similarity 57.1%; Pred. No. 81;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 RPGRKWLGETSSA 14  
 :|||:|||||  
 DB 431 QPGKSWLREESLA 444  
 RESULT 44  
 Q57MD5 SALCH  
 ID Q57MD5 SALCH PRELIMINARY; PRT; 562 AA.  
 AC Q57MD5;  
 DT 10-MAY-2005, integrated into UniProtKB/TrEMBL.  
 DT 10-MAY-2005, sequence version 1.  
 DE Sugar Specific PTS system, fructose-specific transport protein.  
 DE PTS system, fructose-specific IIBC component.  
 GN Name=fruA; OrderedLocNames=SC2220; ORFNames=SCH\_2220;  
 OS Salmonella choleraesuis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=591;  
 RN [1]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=SC-B67;  
 RX PubMed=15781495; DOI=10.1093/nar/gki297;  
 RA Chiu C.-H., Tang P., Chu C., Hu S., Bao Q., Yu J., Chou Y.-Y.,  
 RA Wang H.-S., Lee Y.-S.;  
 RT "The genome sequence of Salmonella enterica serovar Choleraesuis, a  
 RT highly invasive and resistant zoonotic pathogen";  
 RL Nucleic Acids Res. 33:1690-1698(2005).  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL: AE017220; AAX66126.1; -; Genomic DNA.  
 DR GO: 0016021; C:integral to membrane; IEA.  
 DR GO: 0016020; C:membrane; IEA.  
 DR GO: 0008982; F:protein-N(P)-phosphohistidine-sugar phospho. .; IEA.  
 DR GO: 0005351; F:sugar porter activity; IEA.  
 DR GO: 0009401; P:phosphoenolpyruvate-dependent sugar phospho. .; IEA.  
 DR InterPro: IPR013011; PTS\_EIIB\_2.  
 DR InterPro: IPR003352; PTS\_EIIC.  
 DR InterPro: IPR013014; PTS\_EIIC\_2.  
 DR InterPro: IPR003353; PTS\_IIB fruc.  
 DR InterPro: IPR006327; PTS\_IIC fruc.  
 DR Pfam: PF02378; PTS\_EIIC\_1.  
 DR Pfam: PF02379; PTS\_IIB fruc; 1.  
 DR TIGRFAMs: TIGR00829; FRU; 1.  
 DR TIGRFAMs: TIGR01427; PTS\_IIC fruc; 1.  
 DR PROSITE: PS51099; PTS\_EIIB TYPE 2; 1.  
 DR PROSITE: PS51104; PTS\_EIIC TYPE 2; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 562 AA; 57370 MW; 15BA06985E935A74 CRC64;  
 Query Match 54.9%; Score 45; DB 2; Length 562;  
 Best Local Similarity 66.7%; Pred. No. 97;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 3 GKXWLGITSSA 14  
 :|||:|||||  
 DB 59 GKXWLGITGRA 70  
 RESULT 45  
 Q5PE41 SALPA  
 ID Q5PE41 SALPA PRELIMINARY; PRT; 562 AA.  
 AC Q5PE41;  
 DT 04-JAN-2005, integrated into UniProtKB/TrEMBL.  
 DT 04-JAN-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 10.  
 DE PTS system, fructose-specific IIBC component.  
 GN Name=fruA; OrderedLocNames=SPA0647;  
 OS Salmonella paratyphi-a.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=54388;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=ATCC 9150 / SARB42;  
 RX PubMed=15531882; DOI=10.1038/ngl470;  
 RA McClelland M., Sanderson K.E., Clifton S.W., Latreille P.,  
 RA Porwollik S., Sabo A., Meyer R., Bieri T., Ozerky P., McLellan M.,  
 RA Harkins C.R., Wang C., Nguyen C., Berghoff A., Elliott G.,  
 RA Kohlberg S., Strong C., Du F., Carter J., Kremizki C., Layman D.,  
 RA Leonard S., Sun H., Fulton L., Naeh W., Miner T., Minx P.,  
 RA Delehaanty K., Fronick C., Magrini V., Nhan M., Warren W., Florea L.,  
 RA Spieth J., Wilson R.K.;  
 RT "Comparison of genome degradation in Paratyphi A and Typhi, human-  
 RT restricted serovars of Salmonella enterica that cause typhoid";  
 RL Nat. Genet. 36:1268-1274(2004).  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL: CP000026; AAV76647.1; -; Genomic DNA.

DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008982; F:protein-N(Pi)-phosphohistidine-sugar phospho. . .; IEA.  
DR GO; GO:0005351; F:sugar porter activity; IEA.  
DR GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. . .; IEA.  
DR InterPro; IPR013011; PTS\_EIIB\_2.  
DR InterPro; IPR003352; PTS\_EIIC.  
DR InterPro; IPR013014; PTS\_EIIC\_2.  
DR InterPro; IPR003353; PTS\_IIB fruc.  
DR InterPro; IPR006327; PTS\_IIC fruc.  
DR Pfam; PF02378; PTS\_EIIC; 1.  
DR Pfam; PF02379; PTS\_IIB fruc; 1.  
DR TIGRFAMs; TIGR00829; FRU; 1.  
DR TIGRFAMs; TIGR01427; PTS\_IIC fructo; 1.  
DR PROSITE; PS1099; PTS\_EIIB fruc; 1.  
DR PROSITE; PS1104; PTS\_EIIC fruc; 1.  
KW Complete proteome.  
SQ SEQUENCE 562 AA; 57289 MW; B0302A85DDDB865 CRC64;  
  
Query Match 54.9%; Score 45; DB 2; Length 562;  
Best Local Similarity 66.7%; Pred. No. 97;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
QY 3 GKKVWLGETSSA 14  
|||||:|  
DB 59 GKKVWLGDIGRA 70  
  
RESULT 46  
Q8ZNK5 SALTY PRELIMINARY; PRT; 562 AA.  
AC Q8ZNK5;  
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.  
DT 01-MAR-2002, sequence version 1.  
DT 07-FEB-2006, entry version 20.  
DE Fructose-specific transport protein (EC 2.7.1.69).  
GN Name=fruA; OrderedLocusNames=STM2204;  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602;  
[1]  
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RP STRAIN=LT2 / SGSC3412 / ATCC 700720;  
RC MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;  
RX McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;  
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";  
RL Nature 413:852-856 (2001).  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; A500798; AAL21108.1; -; Genomic\_DNA.  
DR BioCyc; STVP9287:STM2204-MONOMER; -.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008982; F:protein-N(Pi)-phosphohistidine-sugar phospho. . .; IEA.  
DR GO; GO:0005351; F:sugar porter activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. . .; IEA.  
DR InterPro; IPR013011; PTS\_EIIB\_2.  
DR InterPro; IPR003352; PTS\_EIIC.  
DR InterPro; IPR013014; PTS\_EIIC\_2.  
DR InterPro; IPR003353; PTS\_IIB fruc.  
DR InterPro; IPR006327; PTS\_IIC fruc.  
DR Pfam; PF02378; PTS\_EIIC; 1.  
DR Pfam; PF02379; PTS\_IIB fruc; 1.  
DR TIGRFAMs; TIGR00829; FRU; 1.

DR TIGRFAMs; TIGR01427; PTS\_IIC fructo; 1.  
DR PROSITE; PS1099; PTS\_EIIB fruc; 1.  
DR PROSITE; PS1104; PTS\_EIIC fruc; 1.  
KW Complete proteome; Transferrase.  
SQ SEQUENCE 562 AA; 57375 MW; 3D6EE64663F8AA2C CRC64;  
  
Query Match 54.9%; Score 45; DB 2; Length 562;  
Best Local Similarity 66.7%; Pred. No. 97;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
QY 3 GKKVWLGETSSA 14  
|||||:|  
DB 59 GKKVWLGDIGRA 70  
  
RESULT 47  
Q8Z592 SALTI PRELIMINARY; PRT; 562 AA.  
AC Q8Z592; Q7CB39;  
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.  
DT 01-MAR-2002, sequence version 1.  
DT 07-MAR-2006, entry version 24.  
DE PTS system, fructose-specific IIBC component (Fructose-specific IIBC component of PTS system).  
GN Name=fruA; OrderedLocusNames=STY2439, t0651;  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=601;  
[1]  
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RP STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J., Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M., Baker S., Basham D., Brooks K., Chillingworth T., Conerton P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.;  
RA "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18.";  
RL Nature 413:848-852 (2001).  
[2]  
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RP STRAIN=Ty2 / ATCC 700931;  
RC MEDLINE=22531367; PubMed=12644504;  
RX DOI=10.1128/JB.185.7.2330-2337.2003;  
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J., Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;  
RA "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2 and CT18.";  
RL J. Bacteriol. 185:2330-2337 (2003).  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; AL627273; CAD02586.1; -; Genomic\_DNA.  
DR EMBL; AE014613; AAO68352.1; -; Genomic\_DNA.  
DR GenomeReviews; AL513382 GR; STV2439.  
DR BioCyc; STV2439:STV2439-MONOMER; -.  
DR BioCyc; STV2439:STV2439-MONOMER; -.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008982; F:protein-N(Pi)-phosphohistidine-sugar phospho. . .; IEA.  
DR GO; GO:0005351; F:sugar porter activity; IEA.  
DR GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. . .; IEA.  
DR InterPro; IPR013011; PTS\_EIIB\_2.  
DR InterPro; IPR003352; PTS\_EIIC.  
DR InterPro; IPR013014; PTS\_EIIC\_2.  
DR InterPro; IPR003353; PTS\_IIB fruc.

```
DR InterPro; IPR006327; PTS_IIC_fruc.
DR Pfam; PF02378; PTS_IIC_1.
DR Pfam; PF02379; PTS_IIB_fruc; 1.
DR TIGRPFAMs; TIGR00629; FRU; 1.
DR TIGRPFAMs; TIGR01427; PTS_IIC_fructo; 1.
DR PROSITE; PSS1099; PTS_IIB_TYPE_2; 1.
DR PROSITE; PSS1104; PTS_IIC_TYPE_2; 1.
DR Complete proteome.
KW SEQUENCE 562 AA; 57289 MW; B0302A85DDDB865 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 562;
Best Local Similarity 66.7%; Pred. No. 97;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 GKXWLGTSAA 14
DB 59 GKXWLGIGRA 70

RESULT 48
Q5SE54_DICDI PRELIMINARY; PRT; 962 AA.
AC Q5SE54;
DT 24-MAY-2005, integrated into UniProtKB/TrEMBL.
DT 24-MAY-2005, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE Hypothetical protein.
OS ORFNames=DD80190225;
GN Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RX PubMed=15875012; DOI=10.1038/nature03481;
RA Eichinger L., Pachbat J.A., Gloeckner G., Rajandream M.A.,
RA Sugang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA Bankal A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren E.J., Davis P.,
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.M.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D.M., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA Hauser H., James K.D., Quiles M., Madan Babu M., Saito T.,
RA Buchrieser C., Wardroper A., Felder M., Thangavelu M., Johnson D.,
RA Knights A., Loulseghe H., Mungall K.L., Oliver K., Price C.,
RA Quail M.A., Urushihara H., Hernandez J., Rabinowitz E., Steffen D.,
RA Sanders M., Ma J., Kohara Y., Sharp S., Simmonds M.N., Spiegler S.,
RA Tivey A., Sugano S., White B., Walker D., Woodward J.R., Winckler T.,
RA Tanaka Y., Shaulsky G., Schleicher M., Weinstock G.M., Rosenthal A.,
RA Cox E.C., Chisholm R.L., Gibbs R.A., Loomis W.F., Platzer M.,
RA Kay R.C., Williams J.G., Dear P.H., Noegel A.A., Barrell B.G.,
RA Kuspa A.;
RT "The genome of the social amoeba Dictyostelium discoideum.";
RL Nature 435:43-57(2005).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; AAFI01000011; EAL72044.1; -; Genomic_DNA.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PD0000018; WD40; 3.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PSS0082; WD_REPEATS_2; 4.
DR PROSITE; PSS0294; WD_REPEATS_REGION; 2.
KW Hypothetical protein; Repeat; WD repeat.

SQ SEQUENCE 962 AA; 105955 MW; 1545D71B703CID96 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 962;
Best Local Similarity 53.8%; Pred. No. 17e+02;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 GKXWLGTSAY 15
DB 711 GKXWLGQTGNIF 723

RESULT 49
Q3HF42_TRIER PRELIMINARY; PRT; 1869 AA.
AC Q3HF42;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Alpha-2-macroglobulin, N-terminal:2-oxo acid dehydrogenase, lipoyl-
DE binding site.
GN ORFNames=IERYDRAFT 2818;
OS Trichodesmium erythraeum IMS101.
OC Bacteria; Cyanobacteria; Oscillatoriales; Trichodesmium.
OX NCBI_TaxID=203124;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IMS101;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RA "Sequencing of the draft genome and assembly of Trichodesmium
RT erythraeum IMS101.";
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IMS101;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RA "Annotation of the draft genome assembly of Trichodesmium erythraeum
RT IMS101.";
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RA Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; AABK04000004; EAO28072.1; -; Genomic_DNA.
GO; GO:0004866; F:endopeptidase inhibitor activity; IEA.
DR InterPro; IPR002890; A2M_N.
DR InterPro; IPR011625; A2M_N_2.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01835; A2M_N; 1.
DR Pfam; PF07703; A2M_N_2; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SEQUENCE 1869 AA; 209056 MW; 9733B25719B73599 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 1869;
Best Local Similarity 60.0%; Pred. No. 3.5e+02;
Matches 9; Conservative 3; Mismatches 1; Indels 2; Gaps 1;

QY 1 RPKXWLGTSAY 15
DB 598 QPKXWLG--TCAAY 610
```

```
RESULT 50
Q36TJ3_MARHY
ID Q36TJ3_MARHY PRELIMINARY; PRT; 197 AA.
AC Q36TJ3;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Phosphonate metabolism.
GN ORFNames=MaquDRAFT_2498;
OS Marinobacter aquaeolei VT8.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Marinobacter.
OX NCBI_TaxID=351348;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=VT8;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Marinobacter aquaeolei
RT VT8.";
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=VT8;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome of Marinobacter aquaeolei VT8.";
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs License
CC -----
DR EMBL; AALG01000003; EAP00158.1; -; Genomic DNA.
DR GO; GO:0015716; P:phosphonate transport; IEA.
SQ SEQUENCE 197 AA; 21341 MW; 8BA55E15F3B7C259 CRC64;

Query Match 53.7%; Score 44; DB 2; Length 197;
Best Local Similarity 53.8%; Pred. No. 47;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 PGKKVWLGETSSA 14
Db 42 PGESVWLADTDGA 54
```

Search completed: June 5, 2006, 12:51:58  
Job time : 175.507 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:31:47 ; Search time 104.384 Seconds

(without alignments)  
65.702 Million cell updates/sec

Title: US-10-645-659A-8

Perfect score: 82

Sequence: 1 SWELGNPNFLKKA 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

9: Geneseqp2005s:\*

10: Geneseqp2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	15	8	ADR88214 Human hep
2	82	100.0	15	8	ADT78181 Functiona
3	82	100.0	15	9	ADU70829 Human hep
4	82	100.0	15	9	AEA42430 Human hep
5	82	100.0	386	8	ADR88207 Human mat
6	82	100.0	386	8	ADT78174 45kDa sub
7	82	100.0	386	9	ADY27057 Heparanas
8	82	100.0	386	9	ADZ18995 Human hep
9	82	100.0	386	9	AEA42423 Human mat
10	82	100.0	460	9	ADY27061 Heparanas
11	82	100.0	486	9	AE887589 Human hep
12	82	100.0	492	9	ADZ18996 Hep106 co
13	82	100.0	493	9	AE887562 Human hep
14	82	100.0	495	9	ADZ18999 Hep109 co
15	82	100.0	497	9	AE887587 Human hep
16	82	100.0	501	9	ADZ19000 HepGS3 co
17	82	100.0	507	9	ADZ19005 HepGS6 co
18	82	100.0	508	9	ADY27058 Human ina
19	82	100.0	526	9	ADZ19006 HepHyalur
20	82	100.0	527	5	ABW07815 Chicken s
21	82	100.0	527	7	ABW02018 Chimeric
22	82	100.0	527	8	ADO63826 Chimeric
23	82	100.0	527	9	ADZ19004 HepGS4 co

97	68	82.9	15	9	ADU70898	Human hep	170	41	50.0	286	3	AAG48720	Aag48720 Arabidops
98	67	81.7	15	9	ADU71196	Human hep	171	41	50.0	286	3	AAG11224	Aag11224 Arabidops
99	64	78.0	15	9	ADU70961	Human hep	172	41	50.0	291	3	AAG48719	Aag48719 Arabidops
100	62	75.6	15	9	ADU71197	Human hep	173	41	50.0	291	3	AAG11223	Aag11223 Arabidops
101	58	70.7	15	9	ADU71198	Human hep	174	41	50.0	298	3	AAG11222	Aag11222 Arabidops
102	58	70.7	528	5	AAE18327	Human hep	175	41	50.0	298	3	AAG48718	Aag48718 Arabidops
103	58	70.7	538	4	AAE18327	Human hep	176	41	50.0	303	4	ABB63233	Abb63233 Drosophil
104	58	70.7	538	4	AAE18327	Human hep	177	41	50.0	431	4	AAB20603	Aau20603 Human sec
105	58	70.7	592	4	AAE18326	Human hep	178	41	50.0	515	8	ADX92843	Adx92843 Plant ful
106	58	70.7	592	4	AAE18326	Human hep	179	41	50.0	538	6	ABM68256	Abm68256 Photorhab
107	58	70.7	592	4	AAE18326	Human hep	180	41	50.0	648	7	ADD45479	Ad45479 Rat Prote
108	58	70.7	592	4	AAE18326	Human hep	181	41	50.0	648	7	AD57446	Ad57446 Rat Prote
109	54	65.9	9	9	ADU70758	Human hep	182	41	50.0	648	9	ABE18192	Aeb18192 Norway ra
110	54	65.9	9	9	ADU70693	Human hep	183	41	50.0	1146	4	AAE00961	Aae00961 Chicken l
111	54	65.9	15	9	ADU70921	Human hep	184	41	50.0	1194	6	ABU18653	Abu18653 Protein e
112	52	63.4	15	9	ADU71044	Human hep	185	40	48.8	45	9	AEC20423	Aec20423 FIV trime
113	52	63.4	651	9	ABE18194	Dog beta-	186	40	48.8	145	6	ABO44374	Ab044374 Corn ster
114	50	61.0	651	9	ABE18193	Cat beta-	187	40	48.8	237	9	ADY53112	Ady53112 Acropora
115	49	59.8	9	9	ADU70795	Human hep	188	40	48.8	237	9	ADY53113	Ady53113 Acropora
116	49	59.8	9	9	ADU70543	Human hep	189	40	48.8	340	4	ABE71784	Abb71784 Drosophil
117	48	58.5	9	9	ADU70759	Human hep	190	40	48.8	433	3	AAG34745	Aag34745 Arabidops
118	48	58.5	9	9	ADU70387	Human hep	191	40	48.8	439	6	ABM67775	Abm67775 Photorhab
119	48	58.5	9	9	ADU70577	Human hep	192	40	48.8	498	9	ADX39546	Adx39546 HIV Gag p
120	48	58.5	15	9	ADU71195	Human hep	193	40	48.8	505	2	AAK48206	Aar48206 A.thalian
121	46	56.1	9	9	ADU70600	Human hep	194	40	48.8	505	3	AAG34744	Aag34744 Arabidops
122	46	56.1	17	9	ADZ18993	Synthetic	195	40	48.8	505	9	AEC59834	Aec59834 Cinnamate
123	45	54.9	463	8	ADZ25621	Aspergill	196	40	48.8	505	10	AEF28197	Aef28197 Lead Cere
124	45	54.9	488	4	AAE181469	Amino aci	197	40	48.8	533	3	AAG34743	Aag34743 Arabidops
125	45	54.9	488	4	AAE181472	Amino aci	198	40	48.8	533	3	AAG34743	Aag34743 Arabidops
126	45	54.9	488	4	AAE181470	Amino aci	199	40	48.8	572	8	ABM84769	Abm84769 Human dia
127	45	54.9	648	4	AAE02444	Murine be	200	40	48.8	572	8	ABM84768	Abm84768 Human dia
128	45	54.9	648	7	ADF28924	Murine be	201	40	48.8	589	8	ABM84772	Abm84772 Human dia
129	45	54.9	648	9	ABE18191	Mouse bet	202	40	48.8	600	7	ADG76363	Adg76363 Human Inc
130	44	53.7	90	5	ABP26310	Streptoco	203	40	48.8	600	8	ABM84771	Abm84771 Human dia
131	44	53.7	364	4	AAE183810	Peptide #	204	40	48.8	613	2	AAW93823	Aaw93823 E. coli G
132	44	53.7	364	4	ABP32748	Peptide #	205	40	48.8	613	2	AAW93828	Aaw93828 Human GUS
133	44	53.7	364	4	ABP32748	Peptide #	206	40	48.8	613	3	AAE28407	Aeb28407 Escherich
134	44	53.7	364	4	ABP27585	Human pep	207	40	48.8	628	9	ABE18189	Aeb18189 Nematode
135	44	53.7	364	4	ABP18230	Protein #	208	40	48.8	633	4	ABE62276	Aeb62276 Mutant he
136	44	53.7	364	4	AAE65943	Human bon	209	40	48.8	633	4	ABE62271	Aeb62271 Heavy cha
137	44	53.7	364	4	AAE53560	Human bra	210	40	48.8	640	2	AAE27210	Aay27210 Amino aci
138	44	53.7	364	4	ABG47603	Human liv	211	40	48.8	648	9	ABE18195	Aeb18195 African g
139	44	53.7	364	4	AAE01554	Peptide #	212	40	48.8	651	4	AAE02443	Aae02443 Human bet
140	44	53.7	364	5	ABG35580	Human pep	213	40	48.8	651	7	ADD45481	Add45481 Human pro
141	44	53.7	377	4	ABU52692	Human dif	214	40	48.8	651	7	ADE57448	Ade57448 Human pro
142	44	53.7	383	4	ABG04768	Novel hum	215	40	48.8	651	8	ADP12392	Adp12392 Protein e
143	44	53.7	438	4	ABU52691	Different	216	40	48.8	651	8	ADQ89808	Adq89808 Antagonis
144	44	53.7	438	9	ADO44021	Amino aci	217	40	48.8	651	9	ADX06032	Adx06032 Cyclin-de
145	44	53.7	438	9	ABE11877	Testis sp	218	40	48.8	651	9	ABE18196	Aeb18196 Human bet
146	44	53.7	463	3	AAE41709	Human ORF	219	40	48.8	651	9	AED74652	Aed74652 Human pla
147	42	51.2	67	7	ADC95420	E. faeciu	220	40	48.8	716	8	ADS44698	Ads44698 Bacterial
148	42	51.2	121	5	ABP02395	Human ORF	221	40	48.8	716	10	AEF94558	Aef94558 Bacillus
149	42	51.2	207	4	ABG03263	Novel hum	222	40	48.8	722	6	AAE33322	Aae33322 L. mexica
150	42	51.2	250	4	ABG79652	Corynebac	223	40	48.8	722	6	ABE33322	Aeb33322 L. mexica
151	42	51.2	250	4	AAE04864	Corynebac	224	40	48.8	722	8	ADJ58609	Adj58609 Fusion pr
152	42	51.2	250	4	AAU71880	C. glutam	225	40	48.8	722	10	AEF05094	Aef05094 Beta-gluc
153	42	51.2	250	4	AAE91908	C. glutami	226	40	48.8	799	5	ABP73572	Abp73572 Candida a
154	42	51.2	250	5	ABE181986	C. glutam	227	40	48.8	909	2	AAE50092	Aae50092 Humanised
155	42	51.2	250	5	ABP98291	Amino aci	228	40	48.8	7067	6	AAE35265	Aae35265 Human P45
156	42	51.2	250	6	ABP98290	Amino aci	229	39.5	48.2	567	8	ADQ36921	Adq36921 Cell prol
157	42	51.2	656	9	ABJ26836	Thymidyla	230	39.5	48.2	594	4	AAU08489	Aau08489 Human VMG
158	42	51.2	250	6	ABJ26837	Thymidyla	231	39.5	48.2	676	5	AAU76161	Aau76161 Bacillus
159	42	51.2	250	6	ABR55231	Amino aci	232	39.5	48.2	676	10	AEF68907	Aef68907 B. stearo
160	42	51.2	250	6	ABR55232	Amino aci	233	39	47.6	52	4	AAE82471	Aam82471 Human imm
161	42	51.2	250	6	ABR55233	Amino aci	234	39	47.6	94	5	ABG93137	Abg93137 S. cerevi
162	42	51.2	656	9	AED25678	Sunflower	235	39	47.6	137	8	ADS16684	Ads16684 Bartonell
163	42	51.2	1420	5	ABP73769	Candida a	236	39	47.6	258	7	ADH87709	Adh87709 Enterococ
164	41	50.0	9	9	ADU70491	Human hep	237	39	47.6	280	7	ADJ80150	Adj80150 Novel hum
165	41	50.0	15	9	ADU71109	Human hep	238	39	47.6	280	9	AEC20252	Aec20252 Human nuc
166	41	50.0	15	9	ADU71042	Human hep	239	39	47.6	354	4	ABE60901	Ab60901 Drosophil
167	41	50.0	121	9	ABE39609	L. pneumo	240	39	47.6	357	2	AAE77394	Aar77394 Fragment
168	41	50.0	125	9	ABE36186	L. pneumo	241	39	47.6	362	8	ADS23337	Ads23337 Bacterial
169	41	50.0	202	7	ABO84289	Pseudomon	242	39	47.6	373	7	ADJ38293	Adj38293 A. cellul



243	39	47.6	375	6	ABP73019	Abp73019 Amino aci	316	38	46.3	410	3	AAG47939	Aag47939 Arabidops
244	39	47.6	389	5	AAM49427	Aam49427 Penicilli	317	38	46.3	431	5	ABB90809	Abb90809 Herbicida
245	39	47.6	403	2	ADC32901	Adc32901 Human nov	318	38	46.3	432	3	AAG47938	Aag47938 Arabidops
246	39	47.6	435	2	AAV19792	B. burgdo	319	38	46.3	451	3	AAG20843	Aag20843 Arabidops
247	39	47.6	454	2	AAV19791	B. burgdo	320	38	46.3	455	3	AAG20842	Aag20842 Arabidops
248	39	47.6	473	8	AD844165	Ad844165 Bacterial	321	38	46.3	488	4	AAB31471	Aab31471 Amino aci
249	39	47.6	478	3	AG29660	Ag29660 Arabidops	322	38	46.3	634	6	ABU18213	Abu18213 Protein e
250	39	47.6	481	3	AG29659	Ag29659 Arabidops	323	38	46.3	726	6	ABR53854	AbR53854 Protein e
251	39	47.6	491	6	ABU20159	Abu20159 Protein e	324	38	46.3	726	7	ADK64928	Adk64928 Disease t
252	39	47.6	620	8	ADS25748	AdS25748 Bacterial	325	38	46.3	2307	4	ABB65433	Abb65433 Drosophil
253	39	47.6	623	8	ADS25902	AdS25902 Bacterial	326	38	46.3	3290	6	ADA34199	Ada34199 Acinetoba
254	39	47.6	647	8	ABU40705	Abu40705 Protein e	327	37.5	45.7	328	6	ABU20472	Abu20472 Protein e
255	39	47.6	649	8	ADS22340	AdS22340 Bacterial	328	37	45.1	38	4	ABG00367	AbG00367 Human nov
256	39	47.6	663	7	AAW49872	Aaw49872 Thermotog	329	37	45.1	55	4	AAU20674	Aau20674 Human nov
257	39	47.6	663	7	ADC26952	AdC26952 Thermotog	330	37	45.1	78	2	AAV35853	Aav35853 Chlamydia
258	39	47.6	663	7	ADF04165	Adf04165 Bacterial	331	37	45.1	96	5	ABP42516	Abp42516 Human ova
259	39	47.6	663	7	AD893948	Ad893948 T. mariti	332	37	45.1	99	4	AAM84777	Aam84777 Human imm
260	39	47.6	666	2	AAW34992	Aaw34992 Thermotog	333	37	45.1	103	8	ABG16641	Abg16641 Novel hum
261	39	47.6	680	2	AAW34564	Aaw34564 Thermotog	334	37	45.1	112	4	ABG16641	Abg16641 Novel hum
262	39	47.6	680	2	AAW35005	Aaw35005 Thermotog	335	37	45.1	156	6	ABU29335	Abu29335 Protein e
263	39	47.6	680	2	AAW49868	Aaw49868 Thermotog	336	37	45.1	157	4	ABG44397	Abg44397 Amino aci
264	39	47.6	680	7	ADC26916	AdC26916 Thermotog	337	37	45.1	158	7	ADH88103	Adh88103 Enterococ
265	39	47.6	680	7	AD893812	Ad893812 T. mariti	338	37	45.1	174	8	ADP84553	Adp84553 Human bre
266	39	47.6	699	3	AAO34445	Aao34445 Candida a	339	37	45.1	192	4	ABB65518	Abb65518 Drosophil
267	39	47.6	699	5	ABP73798	Abp73798 Candida a	340	37	45.1	224	2	AAV37471	Aav37471 Protein i
268	39	47.6	762	6	ABP73022	Abp73022 Amino aci	341	37	45.1	225	3	AGI3478	AgI3478 Arabidops
269	39	47.6	762	7	ADJ38291	Adj38291 A. celluli	342	37	45.1	247	3	AGI3477	AgI3477 Arabidops
270	39	47.6	941	2	AAK07478	Aar07478 Cellulase	343	37	45.1	253	3	AAG43712	Aag43712 Arabidops
271	39	47.6	941	2	AAK77395	Aar77395 Full leng	344	37	45.1	257	3	AAG43711	Aag43711 Arabidops
272	39	47.6	978	4	ABG13924	Abg13924 Novel hum	345	37	45.1	279	3	AAG43710	Aag43710 Arabidops
273	38	46.3	33	6	ABP79666	Abp79666 N. gonorr	346	37	45.1	285	6	ADA35177	Ada35177 Acinetoba
274	38	46.3	65	5	ADH32561	Adh32561 Yeast smo	347	37	45.1	316	9	ABE40822	AbE40822 L. pneumo
275	38	46.3	97	4	ABG10472	Abg10472 Novel hum	348	37	45.1	319	9	ABE37510	AbE37510 L. pneumo
276	38	46.3	119	8	ADH80700	Adh80700 Human pol	349	37	45.1	327	9	ABE48805	AbE48805 Streptomy
277	38	46.3	120	4	AAU14382	Aau14382 Human nov	350	37	45.1	345	6	ABU35760	AbU35760 Protein e
278	38	46.3	120	10	ABE24056	Aee24056 Novel hum	351	37	45.1	374	9	ABM96740	Abm96740 M. xanthu
279	38	46.3	128	8	ADT59176	Adt59176 Plant pol	352	37	45.1	377	6	ABP76792	Abp76792 N. gonorr
280	38	46.3	137	5	ABA47884	Aba47884 p15 Ink4b	353	37	45.1	377	6	ABP81027	Abp81027 N. gonorr
281	38	46.3	138	2	AAK85115	Aar85115 Cell-cycl	354	37	45.1	395	8	ADS29563	AdS29563 Bacterial
282	38	46.3	138	3	AAV88361	Aav88361 Human cel	355	37	45.1	408	7	ABO66862	AbO66862 Klebsiell
283	38	46.3	138	5	AAU74701	Aau74701 Human cel	356	37	45.1	415	4	AAE00417	Aae00417 Lycopersi
284	38	46.3	160	8	ADV89860	Adv89860 Streptoco	357	37	45.1	420	7	ADG62943	AdG62943 Rat Prote
285	38	46.3	160	8	ADV82606	Adv82606 Streptoco	358	37	45.1	420	7	ADG62949	AdG62949 Rat Prote
286	38	46.3	160	8	ADV83250	Adv83250 Streptoco	359	37	45.1	420	7	ADG63970	AdG63970 Disease t
287	38	46.3	160	8	ADV81113	Adv81113 Streptoco	360	37	45.1	420	7	ADG62946	AdG62946 Rat Prote
288	38	46.3	160	8	ADV82665	Adv82665 Streptoco	361	37	45.1	420	7	ADG62946	AdG62946 Rat Prote
289	38	46.3	196	8	ADX89925	Adx89925 Plant ful	362	37	45.1	430	7	ADG63970	AdG63970 Disease t
290	38	46.3	207	8	ADX94844	Adx94844 Plant ful	363	37	45.1	430	8	ADN19013	Adn19013 Bacterial
291	38	46.3	213	8	ADS26087	AdS26087 Bacterial	364	37	45.1	437	7	ADB80185	AdB80185 Mycobacte
292	38	46.3	215	8	ADS25518	AdS25518 Bacterial	365	37	45.1	442	6	ABU38438	Abu38438 Protein e
293	38	46.3	215	8	ADS22628	AdS22628 Bacterial	366	37	45.1	458	8	ADK48652	Adk48652 Streptoco
294	38	46.3	215	8	ADS25199	AdS25199 Bacterial	367	37	45.1	463	8	ADR94871	Adr94871 Novel S.
295	38	46.3	241	8	ADX93484	Adx93484 Plant ful	368	37	45.1	463	9	AEA58741	Aea58741 Streptoco
296	38	46.3	247	2	AAK66291	Aar66291 Mycobacte	369	37	45.1	479	8	ADT56538	Adt56538 Plant pol
297	38	46.3	247	2	AAK63899	Aar63899 M. bovis	370	37	45.1	485	7	ABO77109	AbO77109 Pseudomon
298	38	46.3	247	2	AAW40809	Aaw40809 M. bovis	371	37	45.1	518	6	ABU34324	Abu34324 Protein e
299	38	46.3	247	6	ABU36643	Abu36643 Protein e	372	37	45.1	521	8	ADX92593	AdX92593 Plant ful
300	38	46.3	247	6	ABU34433	Abu34433 Protein e	373	37	45.1	544	2	AAK82213	Aar82213 Talaromyc
301	38	46.3	250	8	ADX76897	Adx76897 Plant ful	374	37	45.1	552	8	ADT57791	Adt57791 Plant pol
302	38	46.3	260	8	ADX88830	Adx88830 Plant ful	375	37	45.1	554	8	ABM84754	Abm84754 Human dia
303	38	46.3	260	8	ADX92376	Adx92376 Plant ful	376	37	45.1	557	8	ADK75975	Adk75975 Plant ful
304	38	46.3	324	3	AAG47940	Aag47940 Arabidops	377	37	45.1	564	6	AAE38212	Aae38212 Human enz
305	38	46.3	328	4	ABB61517	Abb61517 Drosophil	378	37	45.1	582	8	ADY24823	AdY24823 Plant ful
306	38	46.3	344	4	AAK79272	Aak79272 Corynebac	379	37	45.1	585	2	AAV13375	Aav13375 Amino aci
307	38	46.3	344	4	AAK92789	Aak92789 C. glutami	380	37	45.1	587	8	ADOL3883	Ado13883 P. amagaz
308	38	46.3	344	6	ABU26122	Abu26122 Protein e	381	37	45.1	604	9	AEA60269	Aea60269 Streptoco
309	38	46.3	357	3	AAG47948	Aag47948 Arabidops	382	37	45.1	604	9	AEA60269	Aea60269 Streptoco
310	38	46.3	358	3	AAG20844	Aag20844 Arabidops	383	37	45.1	610	8	ADH35329	Adh35329 ENZM prot
311	38	46.3	360	8	ADO62025	Ado62025 Transcrip	384	37	45.1	610	8	ABM84753	Abm84753 Human dia
312	38	46.3	364	7	ADH87100	Adh87100 Enterococ	385	37	45.1	612	8	ABM84752	Abm84752 Human dia
313	38	46.3	367	5	ADP65678	Adp65678 Bifidobac	386	37	45.1	615	8	ADY07633	Ady07633 Plant ful
314	38	46.3	379	3	AAG47947	Aag47947 Arabidops	387	37	45.1	616	8	ABM84751	Abm84751 Human dia
315	38	46.3	383	3	AAG47946	Aag47946 Arabidops	388	37	45.1				

389	37	45.1	624	3	RAG30591	Aeg30591	Arabidops	462	37	45.1	654	6	ABU81251	Human	PRO
390	37	45.1	631	2	AAW77896	Aar77896	Bacterial	463	37	45.1	654	6	ABR60048	Human	sec
391	37	45.1	631	2	AAW08970	Aaw08970	Amino aci	464	37	45.1	654	6	ABR67783	Human	sec
392	37	45.1	631	2	RAY51784	Aay51784	H. influe	465	37	45.1	654	6	ABR65171	Human	sec
393	37	45.1	631	2	AAW54128	Aaw54128	H. influe	466	37	45.1	654	6	ABR68393	Human	sec
394	37	45.1	631	3	AAW80366	Aay80366	H. influe	467	37	45.1	654	6	ABR71805	Human	sec
395	37	45.1	631	3	AAW80366	Aay80366	H. influe	467	37	45.1	654	6	ABR71805	Human	sec
396	37	45.1	648	8	ADO13885	Ado13885	P. amagas	468	37	45.1	654	6	ABR85285	Human	PRO
397	37	45.1	654	4	ADC78497	Adc78497	Human PRO	469	37	45.1	654	6	ABU88975	Human	sec
398	37	45.1	654	4	AAE08243	Aab08243	Human PRO	470	37	45.1	654	6	ABU83055	Human	sec
399	37	45.1	654	4	AAE06593	Aae06593	Human PRO	471	37	45.1	654	6	ABU83055	Human	sec
400	37	45.1	654	4	AAU29036	Aau29036	Human PRO	472	37	45.1	654	6	ABU90459	Novel hum	
401	37	45.1	654	4	ABU58412	Abu58412	Human PRO	473	37	45.1	654	6	ABU83970	Human	sec
402	37	45.1	654	6	ABU71621	Abu71621	Human PRO	474	37	45.1	654	6	ABU93621	Novel hum	
403	37	45.1	654	6	ABU87960	Abu87960	Novel hum	475	37	45.1	654	6	ABR64866	Human	sec
404	37	45.1	654	6	ABU84275	Abu84275	Human sec	476	37	45.1	654	6	ABR68698	Human	sec
405	37	45.1	654	6	ABR66149	Abr66149	Human sec	477	37	45.1	654	6	ABO06514	Human	sec
406	37	45.1	654	6	ABR65539	Abr65539	Human sec	478	37	45.1	654	6	ABR99059	Human	sec
407	37	45.1	654	6	ABU99479	Abu99479	Human sec	479	37	45.1	654	6	ABU56943	Human	PRO
408	37	45.1	654	6	ABU82718	Abu82718	Human PRO	480	37	45.1	654	6	ABU64530	Human	sec
409	37	45.1	654	6	ABU89839	Abu89839	Novel hum	481	37	45.1	654	6	ABU85895	Novel hum	
410	37	45.1	654	6	ABU71476	Abu71476	Human PRO	482	37	45.1	654	6	ABU67376	Human	sec
411	37	45.1	654	6	ABR68088	Abr68088	Human sec	483	37	45.1	654	6	ABU82182	Novel hum	
412	37	45.1	654	6	ABU96141	Abu96141	Novel hum	484	37	45.1	654	6	ABU87193	Human	PRO
413	37	45.1	654	6	ABU92572	Abu92572	Human sec	485	37	45.1	654	6	ABU83665	Human	sec
414	37	45.1	654	6	ABO08649	Abo08649	Human sec	486	37	45.1	654	6	ABO08039	Human	PRO
415	37	45.1	654	6	ABO02701	Abo02701	Human sec	487	37	45.1	654	6	ABO14896	Human	sec
416	37	45.1	654	6	ABR74855	Abr74855	Human sec	488	37	45.1	654	6	ABU81750	Novel hum	
417	37	45.1	654	6	ABR94617	Abr94617	Human sec	489	37	45.1	654	6	ABU65914	Novel hum	
418	37	45.1	654	6	ABU85590	Abu85590	Human PRO	490	37	45.1	654	6	ABR59743	Human	sec
419	37	45.1	654	6	ABU98750	Abu98750	Novel hum	491	37	45.1	654	6	ABU93931	Novel hum	
420	37	45.1	654	6	ABU97965	Abu97965	Novel hum	492	37	45.1	654	6	ABU99784	Novel hum	
421	37	45.1	654	6	ABU91671	Abu91671	Novel hum	493	37	45.1	654	6	ABR66454	Human	sec
422	37	45.1	654	6	ABU71922	Abu71922	Human sec	494	37	45.1	654	6	ABR90872	Human	sec
423	37	45.1	654	6	ABU83364	Abu83364	Human PRO	495	37	45.1	654	6	ABU94299	Human	PRO
424	37	45.1	654	6	ABU86205	Abu86205	Human sec	496	37	45.1	654	6	ABU79181	Human	PRO
425	37	45.1	654	6	ABU67418	Abu67418	Human sec	497	37	45.1	654	6	ABU86510	Human	sec
426	37	45.1	654	6	ABU80446	Abu80446	Human PRO	498	37	45.1	654	6	ABU86815	Novel hum	
427	37	45.1	654	6	ABO01805	Abo01805	Novel hum	499	37	45.1	654	6	ABU94604	Human	PRO
428	37	45.1	654	6	ABR99364	Abr99364	Human sec	500	37	45.1	654	6	ABO04531	Human	PRO
429	37	45.1	654	6	ABR98754	Abr98754	Human sec	501	37	45.1	654	6	ABR70280	Human	sec
430	37	45.1	654	6	ABO16277	Abo16277	Human sec	502	37	45.1	654	6	ABU98445	Human	PRO
431	37	45.1	654	6	ABR92177	Abr92177	Human sec	503	37	45.1	654	6	ABR65844	Human	sec
432	37	45.1	654	6	ABO18818	Abo18818	Human sec	504	37	45.1	654	6	ABR64561	Human	sec
433	37	45.1	654	6	ABR78239	Abr78239	Human sec	505	37	45.1	654	6	ABU15632	Protein e	
434	37	45.1	654	6	ABU84975	Abu84975	Novel hum	506	37	45.1	654	6	ABU79486	Human	PRO
435	37	45.1	654	6	ABO00114	Abo00114	Novel hum	507	37	45.1	654	6	ABU92877	Human	sec
436	37	45.1	654	6	ABO11446	Abo11446	Human sec	508	37	45.1	654	6	ABU95836	Human	PRO
437	37	45.1	654	6	ABO02091	Abo02091	Human sec	509	37	45.1	654	6	ABU91056	Novel hum	
438	37	45.1	654	6	ABU54378	Abu54378	Human sec	510	37	45.1	654	6	ABU90149	Novel hum	
439	37	45.1	654	6	ABU88665	Abu88665	Novel hum	511	37	45.1	654	6	ABO09564	Human	sec
440	37	45.1	654	6	ABU83360	Abu83360	Human sec	512	37	45.1	654	6	ABO10836	Human	sec
441	37	45.1	654	6	ABO06161	Abo06161	Novel hum	513	37	45.1	654	6	ABO10836	Human	sec
442	37	45.1	654	6	ABR59197	Abr59197	Human sec	514	37	45.1	654	6	ABR70890	Human	PRO
443	37	45.1	654	6	ABO09259	Abo09259	Human sec	515	37	45.1	654	6	ABU87498	Human	PRO
444	37	45.1	654	6	ABO11141	Abo11141	Human sec	516	37	45.1	654	6	ABU91366	Human	PRO
445	37	45.1	654	6	ABR66759	Abr66759	Human sec	517	37	45.1	654	6	ABR69670	Human	sec
446	37	45.1	654	6	ABO15972	Abo15972	Human sec	518	37	45.1	654	6	ABU80047	Human	PRO
447	37	45.1	654	6	ABO13678	Abo13678	Human sec	519	37	45.1	654	6	ABU69653	Novel hum	
448	37	45.1	654	6	ABO47393	Abo47393	Human sec	520	37	45.1	654	6	ABU93316	Human	PRO
449	37	45.1	654	6	ABU65581	Abu65581	Human sec	521	37	45.1	654	6	ABO09869	Human	sec
450	37	45.1	654	6	ABO07429	Abo07429	Human PRO	522	37	45.1	654	6	ABO08954	Human	sec
451	37	45.1	654	6	ABO03616	Abo03616	Human sec	523	37	45.1	654	6	ABU10522	Human	sec
452	37	45.1	654	6	ABR67064	Abr67064	Human sec	524	37	45.1	654	6	ABU95531	Human	PRO
453	37	45.1	654	6	ABO15667	Abo15667	Human sec	525	37	45.1	654	6	ABU96740	Novel hum	
454	37	45.1	654	6	ABU55948	Abu55948	Human sec	526	37	45.1	654	6	ABR70585	Human	sec
455	37	45.1	654	6	ABU65276	Abu65276	Human PRO	527	37	45.1	654	6	ABO04936	Novel hum	
456	37	45.1	654	6	ABU95221	Abu95221	Novel hum	528	37	45.1	654	6	ABO08344	Human	sec
457	37	45.1	654	6	ABU71124	Abu71124	Human PRO	529	37	45.1	654	6	ABO14835	Human	sec
458	37	45.1	654	6	ABO07734	Abo07734	Human PRO	530	37	45.1	654	6	ABO05551	Human	sec
459	37	45.1	654	6	ABR69975	Abr69975	Human sec	531	37	45.1	654	6	ABR73940	Human	sec
460	37	45.1	654	6	ABR69308	Abr69308	Human sec	532	37	45.1	654	6	ABR95532	Human	sec
461	37	45.1	654	6	ABO01449	Abo01449	Human PRO	533	37	45.1	654	6	ABR80829	Human	sec
	37	45.1	654	6				534	37	45.1	654	6	ABR81134	Human	sec

535	37	45.1	654	6	ABM00830	Human sec	Abm00830	Human sec	608	37	45.1	654	6	ABM28093	Human sec
536	37	45.1	654	6	ABR88432	Human sec	Abm77253	Human sec	609	37	45.1	654	6	ABO32092	Human sec
537	37	45.1	654	6	ABM77253	Human sec	Abm77253	Human sec	610	37	45.1	654	6	ABM15219	Human sec
538	37	45.1	654	6	ABO28737	Human sec	Abm77253	Human sec	611	37	45.1	654	6	ABM06374	Human sec
539	37	45.1	654	6	ABO31482	Human sec	Abm77253	Human sec	612	37	45.1	654	6	ABM04185	Human sec
540	37	45.1	654	6	ABM07899	Human sec	Abm77253	Human sec	613	37	45.1	654	6	ABM22298	Human sec
541	37	45.1	654	6	ABO40379	Human sec	Abm77253	Human sec	614	37	45.1	654	6	ABM07594	Human sec
542	37	45.1	654	6	ABO35804	Human PRO	Abm77253	Human sec	615	37	45.1	654	6	ABO40684	Human sec
543	37	45.1	654	6	ABO43943	Human PRO	Abm77253	Human sec	616	37	45.1	654	6	ABM35331	Human sec
544	37	45.1	654	6	ADA77778	Human PRO	Abm77253	Human sec	617	37	45.1	654	6	ABM33094	Human sec
545	37	45.1	654	6	ABM24738	Human sec	Abm77253	Human sec	618	37	45.1	654	6	ABO52620	Human PRO
546	37	45.1	654	6	ABD29382	Human sec	Abm77253	Human sec	619	37	45.1	654	6	ABO50180	Human sec
547	37	45.1	654	6	ABO03006	Human sec	Abm77253	Human sec	620	37	45.1	654	6	ABU99174	Human sec
548	37	45.1	654	6	ABR90262	Human sec	Abm77253	Human sec	621	37	45.1	654	6	ABO44226	Human sec
549	37	45.1	654	6	ABM17176	Human sec	Abm77253	Human sec	622	37	45.1	654	6	ABO05856	Human sec
550	37	45.1	654	6	ABR94922	Human sec	Abm77253	Human sec	623	37	45.1	654	6	ABO34847	Human PRO
551	37	45.1	654	6	ABR95227	Human sec	Abm77253	Human sec	624	37	45.1	654	6	ABM18396	Human sec
552	37	45.1	654	6	ABR95227	Human sec	Abm77253	Human sec	625	37	45.1	654	6	ADA16213	Human sec
553	37	45.1	654	6	ABR97729	Human sec	Abm77253	Human sec	626	37	45.1	654	6	ABR97424	Human sec
554	37	45.1	654	6	ABR87517	Human sec	Abm77253	Human sec	627	37	45.1	654	6	ABR80524	Human sec
555	37	45.1	654	6	ABM77558	Human sec	Abm77253	Human sec	628	37	45.1	654	6	ABM01135	Human sec
556	37	45.1	654	6	ABM27788	Human sec	Abm77253	Human sec	629	37	45.1	654	6	ABR88737	Human sec
557	37	45.1	654	6	ABM06069	Human sec	Abm77253	Human sec	630	37	45.1	654	6	ABM13389	Human sec
558	37	45.1	654	6	ABM03575	Human sec	Abm77253	Human sec	631	37	45.1	654	6	ABM20773	Human sec
559	37	45.1	654	6	ABM35026	Human sec	Abm77253	Human sec	632	37	45.1	654	6	ABO41904	Human sec
560	37	45.1	654	6	ABM26263	Human sec	Abm77253	Human sec	633	37	45.1	654	6	ABO42514	Human sec
561	37	45.1	654	6	ABO48045	Human sec	Abm77253	Human sec	634	37	45.1	654	6	ABM10034	Human sec
562	37	45.1	654	6	ABR92787	Human sec	Abm77253	Human sec	635	37	45.1	654	6	ABO38549	Human sec
563	37	45.1	654	6	ABO24548	Human sec	Abm77253	Human sec	636	37	45.1	654	6	ABM32789	Human sec
564	37	45.1	654	6	ABM11559	Human sec	Abm77253	Human sec	637	37	45.1	654	6	ABM22603	Human sec
565	37	45.1	654	6	ABM02660	Human sec	Abm77253	Human sec	638	37	45.1	654	6	ABM74814	Human sec
566	37	45.1	654	6	ABM15956	Human sec	Abm77253	Human sec	639	37	45.1	654	6	ADA79570	Human sec
567	37	45.1	654	6	ABO27517	Human sec	Abm77253	Human sec	640	37	45.1	654	6	ABR96204	Human sec
568	37	45.1	654	6	ABM29008	Human sec	Abm77253	Human sec	641	37	45.1	654	6	ABM02355	Human sec
569	37	45.1	654	6	ABM06984	Human sec	Abm77253	Human sec	642	37	45.1	654	6	ABR86297	Human sec
570	37	45.1	654	6	ABM21078	Human sec	Abm77253	Human sec	643	37	45.1	654	6	ABR86602	Human sec
571	37	45.1	654	6	ABM09424	Human sec	Abm77253	Human sec	644	37	45.1	654	6	ABM16566	Human sec
572	37	45.1	654	6	ABO41294	Human sec	Abm77253	Human sec	645	37	45.1	654	6	ABM29618	Human sec
573	37	45.1	654	6	ABO36109	Human PRO	Abm77253	Human sec	646	37	45.1	654	6	ABO29042	Human sec
574	37	45.1	654	6	ABO43638	Human PRO	Abm77253	Human sec	647	37	45.1	654	6	ABM23823	Human sec
575	37	45.1	654	6	ABM76338	Human sec	Abm77253	Human sec	648	37	45.1	654	6	ABM23213	Human sec
576	37	45.1	654	6	ABM76034	Human sec	Abm77253	Human sec	649	37	45.1	654	6	ABM21993	Human sec
577	37	45.1	654	6	ABM25653	Human sec	Abm77253	Human sec	650	37	45.1	654	6	ABO37634	Human sec
578	37	45.1	654	6	ABM25958	Human sec	Abm77253	Human sec	651	37	45.1	654	6	ABM28398	Human sec
579	37	45.1	654	6	ABO03311	Human sec	Abm77253	Human sec	652	37	45.1	654	6	ABM28703	Human sec
580	37	45.1	654	6	ABO02396	Human sec	Abm77253	Human sec	653	37	45.1	654	6	ABM66347	Human sec
581	37	45.1	654	6	ABR90567	Human sec	Abm77253	Human sec	654	37	45.1	654	6	ABM66347	Human sec
582	37	45.1	654	6	ABR73635	Human sec	Abm77253	Human sec	655	37	45.1	654	6	ABM75729	Human sec
583	37	45.1	654	6	ABO16887	Human sec	Abm77253	Human sec	656	37	45.1	654	6	ABM34009	Human sec
584	37	45.1	654	6	ABR94312	Human sec	Abm77253	Human sec	657	37	45.1	654	6	ABM34314	Human sec
585	37	45.1	654	6	ABR75819	Human sec	Abm77253	Human sec	658	37	45.1	654	6	ABO20245	Human sec
586	37	45.1	654	6	ADA18238	Human sec	Abm77253	Human sec	659	37	45.1	654	6	ABO21160	Human sec
587	37	45.1	654	6	ABO32787	Human sec	Abm77253	Human sec	660	37	45.1	654	6	ABO22075	Human sec
588	37	45.1	654	6	ABR71195	Human sec	Abm77253	Human sec	661	37	45.1	654	6	ABR96509	Human sec
589	37	45.1	654	6	ABR93092	Human sec	Abm77253	Human sec	662	37	45.1	654	6	ABR85687	Human sec
590	37	45.1	654	6	ABR93397	Human sec	Abm77253	Human sec	663	37	45.1	654	6	ABR99669	Human sec
591	37	45.1	654	6	ABR87822	Human sec	Abm77253	Human sec	664	37	45.1	654	6	ABM00220	Human sec
592	37	45.1	654	6	ABO27822	Human sec	Abm77253	Human sec	665	37	45.1	654	6	ABM00525	Human sec
593	37	45.1	654	6	ABO29957	Human sec	Abm77253	Human sec	666	37	45.1	654	6	ABO29652	Human sec
594	37	45.1	654	6	ABO33166	Human PRO	Abm77253	Human sec	667	37	45.1	654	6	ABM23518	Human sec
595	37	45.1	654	6	ABM04854	Human sec	Abm77253	Human sec	668	37	45.1	654	6	ABM29313	Human sec
596	37	45.1	654	6	ABM08814	Human sec	Abm77253	Human sec	669	37	45.1	654	6	ABO38244	Human sec
597	37	45.1	654	6	ABO36414	Human sec	Abm77253	Human sec	670	37	45.1	654	6	ABO45544	Human PRO
598	37	45.1	654	6	ABO35499	Human PRO	Abm77253	Human sec	671	37	45.1	654	6	ADA42358	Human sec
599	37	45.1	654	6	ABO39464	Human sec	Abm77253	Human sec	672	37	45.1	654	6	ADA81297	Human sec
600	37	45.1	654	6	ABM10339	Human sec	Abm77253	Human sec	673	37	45.1	654	6	ABO16582	Human sec
601	37	45.1	654	6	ABM11864	Human sec	Abm77253	Human sec	674	37	45.1	654	6	ABO18208	Human sec
602	37	45.1	654	6	ABO52010	Human PRO	Abm77253	Human sec	675	37	45.1	654	6	ABO22635	Human PRO
603	37	45.1	654	6	ABO52315	Human PRO	Abm77253	Human sec	676	37	45.1	654	6	ABO22940	Human PRO
604	37	45.1	654	6	ABO23633	Human sec	Abm77253	Human sec	677	37	45.1	654	6	ABR92482	Human sec
605	37	45.1	654	6	ABR97119	Human sec	Abm77253	Human sec	678	37	45.1	654	6	ABR81439	Human sec
606	37	45.1	654	6	ABR86907	Human sec	Abm77253	Human sec	679	37	45.1	654	6	ABO17525	Human PRO
607	37	45.1	654	6	ABM10949	Human sec	Abm77253	Human sec	680	37	45.1	654	6	ABM77863	Human sec

681	37	45.1	654	6	ABR89652	AbR89652 Human sec	754	37	45.1	654	7	ABO50790	ABO50790 Human sec
682	37	45.1	654	6	ABM26568	AbM26568 Human sec	755	37	45.1	654	7	ABO05246	ABO05246 Human sec
683	37	45.1	654	6	ABM13694	AbM13694 Human sec	756	37	45.1	654	7	ABR74550	ABR74550 Human sec
684	37	45.1	654	6	ABO28432	ABO28432 Human sec	757	37	45.1	654	7	ABR77029	ABR77029 Human sec
685	37	45.1	654	6	ABO30262	ABO30262 Human sec	758	37	45.1	654	7	ADA16637	ADA16637 Human sec
686	37	45.1	654	6	ABM07289	AbM07289 Human sec	759	37	45.1	654	7	ABM17786	ABM17786 Human sec
687	37	45.1	654	6	ABM03880	AbM03880 Human sec	760	37	45.1	654	7	ABR95837	ABR95837 Human sec
688	37	45.1	654	6	ABO37024	ABO37024 Human sec	761	37	45.1	654	7	ADA13066	ADA13066 Human sec
689	37	45.1	654	6	ABO41599	ABO41599 Human sec	762	37	45.1	654	7	ABO21770	ABO21770 Human sec
690	37	45.1	654	6	ABO35194	ABO35194 Human PRO	763	37	45.1	654	7	ABO119940	ABO119940 Human sec
691	37	45.1	654	6	ABM25043	AbM25043 Human sec	764	37	45.1	654	7	ABO24243	ABO24243 Human sec
692	37	45.1	654	6	ABO47435	ABO47435 Human sec	765	37	45.1	654	7	ABR85392	ABR85392 Human sec
693	37	45.1	654	6	ABO47740	ABO47740 Human sec	766	37	45.1	654	7	ABM10644	ABM10644 Human sec
694	37	45.1	654	6	ABO48350	ABO48350 Human sec	767	37	45.1	654	7	ABM76643	ABM76643 Human sec
695	37	45.1	654	6	ABO511400	ABO511400 Human PRO	768	37	45.1	654	7	ABR89347	ABR89347 Human sec
696	37	45.1	654	6	ABO51705	ABO51705 Human PRO	769	37	45.1	654	7	ABM12474	ABM12474 Human sec
697	37	45.1	654	6	ABO50485	ABO50485 Human sec	770	37	45.1	654	7	ABM05764	ABM05764 Human sec
698	37	45.1	654	6	ABR99609	ABR99609 Human sec	771	37	45.1	654	7	ABO34889	ABO34889 Human PRO
699	37	45.1	654	6	ABM16871	AbM16871 Human sec	772	37	45.1	654	7	ABM02965	ABM02965 Human sec
700	37	45.1	654	6	ABO17903	ABO17903 Human sec	773	37	45.1	654	7	ABM18943	ABM18943 Human sec
701	37	45.1	654	6	ABO20855	ABO20855 Human sec	774	37	45.1	654	7	ABM19248	ABM19248 Human sec
702	37	45.1	654	6	ABR96814	ABR96814 Human sec	775	37	45.1	654	7	ABO46459	ABO46459 Human PRO
703	37	45.1	654	6	ABM12169	AbM12169 Human sec	776	37	45.1	654	7	ABO48960	ABO48960 Human sec
704	37	45.1	654	6	ABM16261	AbM16261 Human sec	777	37	45.1	654	7	ADA41934	ADA41934 Human sec
705	37	45.1	654	6	ABM24128	ABM24128 Human sec	778	37	45.1	654	7	ABR69003	ABR69003 Human sec
706	37	45.1	654	6	ABM14609	ABM14609 Human sec	779	37	45.1	654	7	ABR89042	ABR89042 Human sec
707	37	45.1	654	6	ABM04490	ABM04490 Human sec	780	37	45.1	654	7	ABR72415	ABR72415 Human sec
708	37	45.1	654	6	ABM06679	AbM06679 Human sec	781	37	45.1	654	7	ABR74245	ABR74245 Human sec
709	37	45.1	654	6	ABM09119	AbM09119 Human sec	782	37	45.1	654	7	ABO18513	ABO18513 Human sec
710	37	45.1	654	6	ABO39159	ABO39159 Human sec	783	37	45.1	654	7	ADA17281	ADA17281 Human sec
711	37	45.1	654	6	ABM75424	ABM75424 Human sec	784	37	45.1	654	7	ABR80219	ABR80219 Human sec
712	37	45.1	654	6	ABM25348	ABM25348 Human sec	785	37	45.1	654	7	ABM01440	ABM01440 Human sec
713	37	45.1	654	6	ABM19858	ABM19858 Human sec	786	37	45.1	654	7	ABM02050	ABM02050 Human sec
714	37	45.1	654	6	ABO46764	ABO46764 Human PRO	787	37	45.1	654	7	ABR87212	ABR87212 Human sec
715	37	45.1	654	6	ABO47069	ABO47069 Human PRO	788	37	45.1	654	7	ABM12779	ABM12779 Human sec
716	37	45.1	654	6	ADA83095	ADA83095 Human sec	789	37	45.1	654	7	ABM30533	ABM30533 Human sec
717	37	45.1	654	6	ABR71500	ABR71500 Human sec	790	37	45.1	654	7	ABM24433	ABM24433 Human sec
718	37	45.1	654	6	ABR72110	ABR72110 Human sec	791	37	45.1	654	7	ABO29347	ABO29347 Human sec
719	37	45.1	654	6	ABR98449	ABR98449 Human sec	792	37	45.1	654	7	ABO31177	ABO31177 Human sec
720	37	45.1	654	6	ABO06819	ABO06819 Human sec	793	37	45.1	654	7	ABM14304	ABM14304 Human sec
721	37	45.1	654	6	ABR84772	ABR84772 Human sec	794	37	45.1	654	7	ABM09729	ABM09729 Human sec
722	37	45.1	654	6	ABR73330	ABR73330 Human sec	795	37	45.1	654	7	ABO38854	ABO38854 Human sec
723	37	45.1	654	6	ABR76424	ABR76424 Human sec	796	37	45.1	654	7	ABM34619	ABM34619 Human sec
724	37	45.1	654	6	ABR73025	ABR73025 Human sec	797	37	45.1	654	7	ABO51095	ABO51095 Human sec
725	37	45.1	654	6	ABM18091	AbM18091 Human sec	798	37	45.1	654	7	ADA42784	ADA42784 Human sec
726	37	45.1	654	6	ABO20550	ABO20550 Human sec	799	37	45.1	654	7	ABO03921	ABO03921 Human sec
727	37	45.1	654	6	ABO25293	ABO25293 Human PRO	800	37	45.1	654	7	ABO10391	ABO10391 Human PRO
728	37	45.1	654	6	ABO25598	ABO25598 Human PRO	801	37	45.1	654	7	ABR77634	ABR77634 Human sec
729	37	45.1	654	6	ABR94007	ABR94007 Human sec	802	37	45.1	654	7	ABR78844	ABR78844 Human sec
730	37	45.1	654	6	ABR79914	ABR79914 Human sec	803	37	45.1	654	7	ABO23938	ABO23938 Human sec
731	37	45.1	654	6	ABM11254	AbM11254 Human sec	804	37	45.1	654	7	ABR93702	ABR93702 Human sec
732	37	45.1	654	6	ABO32861	ABO32861 Human sec	805	37	45.1	654	7	ABM01745	ABM01745 Human sec
733	37	45.1	654	6	ABO30567	ABO30567 Human sec	806	37	45.1	654	7	ABM78168	ABM78168 Human sec
734	37	45.1	654	6	ABO30872	ABO30872 Human sec	807	37	45.1	654	7	ABR89957	ABR89957 Human sec
735	37	45.1	654	6	ABM27178	AbM27178 Human sec	808	37	45.1	654	7	ABM27483	ABM27483 Human sec
736	37	45.1	654	6	ABM23923	AbM23923 Human sec	809	37	45.1	654	7	ABM13084	ABM13084 Human sec
737	37	45.1	654	6	ABM05459	ABM05459 Human sec	810	37	45.1	654	7	ABO31787	ABO31787 Human sec
738	37	45.1	654	6	ABM15524	AbM15524 Human sec	811	37	45.1	654	7	ABM13999	ABM13999 Human sec
739	37	45.1	654	6	ABM08509	ABM08509 Human sec	812	37	45.1	654	7	ABM08204	ABM08204 Human sec
740	37	45.1	654	6	ABO43209	ABO43209 Human sec	813	37	45.1	654	7	ABO40074	ABO40074 Human sec
741	37	45.1	654	6	ABO37939	ABO37939 Human sec	814	37	45.1	654	7	ABM74509	ABM74509 Human sec
742	37	45.1	654	6	ABO45849	ABO45849 Human PRO	815	37	45.1	654	7	ABM33704	ABM33704 Human sec
743	37	45.1	654	6	ABM66652	ABM66652 Human sec	816	37	45.1	654	7	ABM20163	ABM20163 Human sec
744	37	45.1	654	6	ADB20138	ADB20138 Human sec	817	37	45.1	654	7	ABO48655	ABO48655 Human sec
745	37	45.1	654	6	ABM19553	AbM19553 Human sec	818	37	45.1	654	7	ABR72720	ABR72720 Human sec
746	37	45.1	654	6	ABO49265	ABO49265 Human sec	819	37	45.1	654	7	ABO15362	ABO15362 Human sec
747	37	45.1	654	6	ABO49570	ABO49570 Human sec	820	37	45.1	654	7	ABR85077	ABR85077 Human sec
748	37	45.1	654	6	ADA78390	ADA78390 Human sec	821	37	45.1	654	7	ABO15057	ABO15057 Human sec
749	37	45.1	654	6	ABR88127	ABR88127 Human sec	822	37	45.1	654	7	ABO17192	ABO17192 Human sec
750	37	45.1	654	6	ABM26873	AbM26873 Human sec	823	37	45.1	654	7	ABM17481	ABM17481 Human sec
751	37	45.1	654	6	ABM03270	ABM03270 Human sec	824	37	45.1	654	7	ABR85382	ABR85382 Human sec
752	37	45.1	654	6	ABO39769	ABO39769 Human sec	825	37	45.1	654	7	ABO17586	ABO17586 Human PRO
753	37	45.1	654	7	ABO49875	ABO49875 Human sec	826	37	45.1	654	7	ABM76948	ABM76948 Human sec

827	37	45.1	654	7	ABO28127	Human sec	900	37	45.1	654	8	ADI18659	Human sec
828	37	45.1	654	7	ABM22908	Human sec	901	37	45.1	654	8	ADI165379	Human sec
829	37	45.1	654	7	ABM30228	Human sec	902	37	45.1	654	8	ADI37642	Human sec
830	37	45.1	654	7	ABM21688	Human sec	903	37	45.1	654	8	ADH97438	Human sec
831	37	45.1	654	7	ABM21383	Human sec	904	37	45.1	654	8	ADH165806	Human sec
832	37	45.1	654	7	ABM214914	Human sec	905	37	45.1	654	8	ADH60549	Human sec
833	37	45.1	654	7	ABO40989	Human sec	906	37	45.1	654	8	ADH99606	Human sec
834	37	45.1	654	7	ABO36719	Human sec	907	37	45.1	654	8	ADL08799	Human sec
835	37	45.1	654	7	ABO37329	Human sec	908	37	45.1	654	8	ADJ54664	Human PRO
836	37	45.1	654	7	ABM75119	Human sec	909	37	45.1	654	8	ADM25140	Human sec
837	37	45.1	654	7	ABM33399	Human sec	910	37	45.1	654	8	ADM29890	Human sec
838	37	45.1	654	7	ABO46154	Human PRO	911	37	45.1	654	8	ADJ64435	Human PRO
839	37	45.1	654	7	ADA82461	Human sec	912	37	45.1	654	8	ADK47677	Streptoco
840	37	45.1	654	7	ABM31753	Human sec	913	37	45.1	654	8	ADM31331	Novel hum
841	37	45.1	654	7	ABM31143	Human sec	914	37	45.1	654	8	ADM36378	Novel hum
842	37	45.1	654	7	ADB77703	Human sec	915	37	45.1	654	8	ADM40183	Novel hum
843	37	45.1	654	7	ADB74839	Human sec	916	37	45.1	654	8	ADO06212	Human PRO
844	37	45.1	654	7	ADB85769	Human sec	917	37	45.1	654	8	ADN37791	Novel hum
845	37	45.1	654	7	ABM32058	Human sec	918	37	45.1	654	8	ADR11064	Human sec
846	37	45.1	654	7	ABM32363	Human sec	919	37	45.1	654	8	ADR17973	Human sec
847	37	45.1	654	7	ABM31448	Human sec	920	37	45.1	654	8	ADH74612	Human sec
848	37	45.1	654	7	ABM30838	Human sec	921	37	45.1	654	8	ADT03649	Human sec
849	37	45.1	654	7	ADC28485	Human sec	922	37	45.1	654	9	AEA37887	Human sec
850	37	45.1	654	7	ADC39685	Human sec	923	37	45.1	654	9	AED23921	Human sec
851	37	45.1	654	7	ADC40199	Human sec	924	37	45.1	654	9	AED50021	Novel hum
852	37	45.1	654	7	ADC19023	Human sec	925	37	45.1	654	10	AE669090	Beta-gala
853	37	45.1	654	7	ADC34323	Human sec	926	37	45.1	681	7	ABO80085	Pseudomon
854	37	45.1	654	7	ADC29378	Human sec	927	37	45.1	683	8	ABM84750	Human dia
855	37	45.1	654	7	ADC28909	Human sec	928	37	45.1	722	8	ADO13886	P. amagas
856	37	45.1	654	7	ADC40794	Human sec	929	37	45.1	761	4	ABG02677	Novel hum
857	37	45.1	654	7	ADC19451	Human sec	930	37	45.1	764	4	ABG09702	Novel hum
858	37	45.1	654	7	ADC33899	Human sec	931	37	45.1	780	3	AGG30590	Arabidops
859	37	45.1	654	7	ADC12969	Human sec	932	37	45.1	780	7	ADB95076	A. thalia
860	37	45.1	654	7	ADC12421	Human sec	933	37	45.1	782	4	ABG16643	Novel hum
861	37	45.1	654	7	ADD05499	Human sec	934	37	45.1	786	6	ABU02588	S. pneumo
862	37	45.1	654	7	ADD04976	Human sec	935	37	45.1	803	4	AAU34101	Staphyloc
863	37	45.1	654	7	ADD03982	Human sec	936	37	45.1	804	2	AAW22709	Leucyl-tr
864	37	45.1	654	7	ADD03558	Human sec	937	37	45.1	804	6	ABU15896	Protein e
865	37	45.1	654	7	ADE34810	Human sec	938	37	45.1	804	6	AAE36398	Staphyloc
866	37	45.1	654	7	ADG02494	Novel hum	939	37	45.1	804	9	ADW94769	prolifera
867	37	45.1	654	7	ADG01201	Novel hum	940	37	45.1	805	4	AAU36693	Staphyloc
868	37	45.1	654	7	ADF95376	Novel hum	941	37	45.1	805	6	ADA89537	Staphyloc
869	37	45.1	654	7	ADG12191	Novel hum	942	37	45.1	805	6	ABM71188	Staphyloc
870	37	45.1	654	7	ADH08851	Human PRO	943	37	45.1	834	8	ADSL6669	B. hensel
871	37	45.1	654	7	ADH59293	Human sec	944	37	45.1	843	8	ADM92249	S. pneumon
872	37	45.1	654	7	ADI38072	Human sec	945	37	45.1	844	4	AAU37922	Streptoco
873	37	45.1	654	7	ADJ26340	Human sec	946	37	45.1	857	6	ABU46277	Protein e
874	37	45.1	654	7	ADL32632	Novel hum	947	37	45.1	1059	8	ADK99177	Streptoco
875	37	45.1	654	7	ADM30166	Novel hum	948	37	45.1	1205	9	ABM97591	M. xanthu
876	37	45.1	654	8	ADE79255	Human sec	949	37	45.1	1597	4	ABB60567	Drosophill
877	37	45.1	654	8	ADE79679	Human sec	950	36.5	44.5	72	8	ABO55120	Human gen
878	37	45.1	654	8	ADE79679	Human sec	951	36.5	44.5	87	4	AAO10694	Human pol
879	37	45.1	654	8	ADE73355	Human sec	952	36.5	44.5	116	4	ABG00431	Novel hum
880	37	45.1	654	8	ADE74163	Human sec	953	36.5	44.5	126	8	ADQ36997	Cell prol
881	37	45.1	654	8	ADE73890	Human sec	954	36.5	44.5	126	8	ADQ36997	Rice acre
882	37	45.1	654	8	ADE74775	Human sec	955	36.5	44.5	126	8	ADQ15667	Bacterial
883	37	45.1	654	8	ADE99444	Human sec	956	36.5	44.5	443	8	ADS25494	Bacterial
884	37	45.1	654	8	ADE98563	Human sec	957	36.5	44.5	443	8	ADS25995	Bacterial
885	37	45.1	654	8	ADE98990	Human sec	958	36.5	44.5	443	8	ADS2580	Bacterial
886	37	45.1	654	8	ADG40460	Human sec	959	36.5	44.5	581	4	AAU41714	Propionib
887	37	45.1	654	8	ADF73854	Human sec	960	36	43.9	581	6	ABM38233	Propionib
888	37	45.1	654	8	ADF95988	Novel hum	961	36	43.9	35	6	ABU01244	S. pneumo
889	37	45.1	654	8	ADG04259	Novel hum	962	36	43.9	49	4	ABG01573	Novel hum
890	37	45.1	654	8	ADG00419	Novel hum	963	36	43.9	65	4	AAU65541	Propionib
891	37	45.1	654	8	ADG82675	Human PRO	964	36	43.9	65	6	ABM62060	Propionib
892	37	45.1	654	8	ADG92273	Human sec	965	36	43.9	66	4	AAO03574	Human pol
893	37	45.1	654	8	ADG92700	Human sec	966	36	43.9	75	3	AAAG21635	Arabidops
894	37	45.1	654	8	ADH25956	Novel hum	967	36	43.9	90	4	AAO06864	Human pol
895	37	45.1	654	8	ADH32925	Human PRO	968	36	43.9	106	4	ABG26789	Novel hum
896	37	45.1	654	8	ADH20489	Human sec	969	36	43.9	112	4	AAU07425	Human hep
897	37	45.1	654	8	ADH07344	Human sec	970	36	43.9	126	8	ADH89866	Plant ful
898	37	45.1	654	8	ADH59889	Human sec	971	36	43.9	129	4	AGH82623	S. epider
899	37	45.1	654	8	ADH06917	Human sec	972	36	43.9	135	4	AAO05901	Human pol

973 36 43.9 147 4 ABG15872 Novel hum  
 974 36 43.9 152 4 AAB45807 Zebratish  
 975 36 43.9 152 9 AED08573 Zebratish  
 976 36 43.9 153 4 ABG18652 Novel hum  
 977 36 43.9 193 4 AAM14443 Peptide #  
 978 36 43.9 193 4 AAM15868 Peptide #  
 979 36 43.9 193 4 ABB36188 Peptide #  
 980 36 43.9 193 4 ABB33391 Peptide #  
 981 36 43.9 193 4 AAM28376 Peptide #  
 982 36 43.9 193 4 AAM26856 Peptide #  
 983 36 43.9 193 4 AAM29680 Peptide #  
 984 36 43.9 193 4 ABB28216 Human pep  
 985 36 43.9 193 4 ABB29682 Peptide #  
 986 36 43.9 193 4 ABB18850 Protein #  
 987 36 43.9 193 4 AAM69353 Human bon  
 988 36 43.9 193 4 AAM66570 Human bon  
 989 36 43.9 193 4 AAM55668 Human bra  
 990 36 43.9 193 4 AAM54176 Human bra  
 991 36 43.9 193 4 AAM56966 Human bra  
 992 36 43.9 193 4 ABG48238 Human liv  
 993 36 43.9 193 4 ABG51029 Human liv  
 994 36 43.9 193 4 AAM02170 Peptide #  
 995 36 43.9 193 4 AAM04882 Peptide #  
 996 36 43.9 193 4 ABG36222 Human pep  
 997 36 43.9 193 5 ABG38972 Human pep  
 998 36 43.9 209 7 ADC00990 Enterohae  
 999 36 43.9 213 6 ABU33837 Protein e  
 1000 36 43.9 228 6 ABU19393 Protein e

## ALIGNMENTS

RESULT 1  
 ADR88214  
 ID ADR88214 standard; peptide; 15 AA.

AC ADR88214;

DT 18-NOV-2004 (first entry)

XX Human heparanase epitope pep8.

XX Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;  
 KW autoimmune disorder; cancer; angiogenesis; metastatic disease;  
 KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;  
 KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;  
 KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;  
 KW human; heparanase; enzyme; epitope.

XX Homo sapiens.

OS US2004170631-A1.

PN 02-SEP-2004.

PD 28-NOV-2003; 2003US-00722502.

PF 02-SEP-1997; 97US-00922170.

PR 01-MAY-1998; 98US-00071739.

PR 04-NOV-1998; 98US-00186200.

PR 19-FEB-2003; 2003US-00368044.

PR 22-AUG-2003; 2003US-00645659.

XX (YACO/) YACOBY-ZEEVI O.

PA (PERE/) PERETZ T.

PA (MIRO/) MIRON D.

PA (SHLO/) SHLOMI Y.

PA (PECK/) PECKER I.

PA (AYAL/) AYAL-HERSHKOVITZ M.

PA (FEIN/) FEINSTEIN E.

PA (VDEL/) VAN GELDER J M.

PA (VLOD/) VLODAVSKY I.

PA (FRIE/) FRIEDMANN Y.

XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;

PI Ayal-Herskovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;

PI Friedmann Y;

XX WPI; 2004-625084/60.

DR Targeted drug delivery to a heparanase-expressing tissue of a patient,  
 XX useful for treating heparanase-associated conditions such as inflammation  
 PT or cancer, comprises administering a drug and an anti-heparanase antibody  
 PT complex.

XX Claim 7; SEQ ID NO 8; 58pp; English.

CC The invention relates to a method of targeted drug delivery to a tissue  
 CC of a patient, the tissue expressing heparanase. The method comprises  
 CC providing a complex of a drug directly or indirectly linked to an anti-  
 CC heparanase antibody, and administering the complex to the patient. In the  
 CC targeted drug delivery, the antibody comprises an antibody or its portion  
 CC capable of specifically binding to at least one epitope of a heparanase  
 CC protein. The composition and methods of the invention are useful for  
 CC diagnosing, preventing or treating conditions associated with heparanase  
 CC catalytic activity (e.g. an inflammatory disorder, wound, scar,  
 CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell  
 CC proliferation, invasion of circulating tumour cells and metastatic  
 CC disease), for purifying heparanase, or for developing drugs for those  
 CC heparanase-associated conditions. The vasculopathy is atherosclerosis,  
 CC restenosis or aneurysm. The cancerous condition is a solid cancer or a  
 CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy  
 CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous  
 CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous  
 CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,  
 CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and  
 CC multiple myeloma. The solid cancer is selected from tumours in lip and  
 CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,  
 CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,  
 CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of  
 CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue  
 CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,  
 CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic  
 CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary  
 CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,  
 CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,  
 CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,  
 CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's  
 CC sarcoma. The present sequence is human heparanase epitope.

XX Sequence 15 AA;

Query Match 100.0%; Score 82; DB 8; Length 15;

Best Local Similarity 100.0%; Pred. No. 2.4e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SWELGNENPSFLKKA 15

Db 1 SWELGNENPSFLKKA 15

RESULT 2

ADT78181

ID ADT78181 standard; peptide; 15 AA.

XX ADT78181;

XX 13-JAN-2005 (first entry)

XX Functional peptide epitope of human heparanase, pep8.

XX Antibody; epitope; heparanase; pathological condition; angiogenesis;

KW cell proliferation; cancerous condition; tumour cell invasion;

KW metastatic disease; heparanase-related disorder; inflammatory disorder;

KW wound; scar; vasculopathy; autoimmune condition; renal disease;

KW cytostatic; antiinflammatory; vulnery; antiarteriosclerotic;  
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.  
 XX

OS Homo sapiens.

PN US2004213789-A1.

PD 28-OCT-2004.

PF 22-AUG-2003; 2003US-00645659.

XX 02-SEP-1997; 97US-00922170.

PR 01-MAY-1998; 98US-00071739.

PR 04-NOV-1998; 98US-00186200.

PR 19-FEB-2003; 2003US-00368044.

XX (YACO/) YACOBY-ZEEVI O.

PA (PERE/) PERETZ T.

PA (MIRO/) MIRON D.

PA (SHLO/) SHLOMI Y.

PA (PECK/) PECKER I.

PA (AYAL/) AYAL-HERSHKOVITZ M.

PA (FEIN/) FEINSTEIN E.

PA (GELD/) GELDER J M V.

PA (VLOD/) VLODAVSKY I.

PA (FRIE/) FRIEDMANN Y.

XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;

PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;

PI Friedmann Y;

XX WPI; 2004-774790/76.

XX New neutralizing monoclonal anti-heparanase antibodies, useful for

PT detecting, treating or preventing cancer, inflammatory or autoimmune

PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.

XX Claim 67; SEQ ID NO 8; 69pp; English.

PS The invention relates to an isolated antibody or antibody portion capable

CC of specifically binding to or elicited by at least one epitope of a

CC heparanase protein, where the heparanase protein is at least 60%

CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and

CC where at least one epitope comprises a sequence at least 70% homologous

CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)

CC a hybridoma cell line comprising a cell line for producing the monoclonal

CC antibody, (b) a method for detecting, treating or preventing a

CC pathological condition or a heparanase-related disorder or condition in a

CC subject, (c) a method for monitoring the state of a heparanase-related

CC disorder or condition in a subject, and (d) a pharmaceutical composition

CC comprising the isolated anti-heparanase antibody or antibody portion and

CC a pharmaceutical carrier. The antibody, methods, and composition are

CC useful for detecting, treating, preventing or monitoring a pathological

CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition

CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,

CC or prostate cancer), minor cell proliferation, invasion of circulating  
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder  
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy  
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or  
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,  
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell  
 CC carcinoma) in a mammal. This sequence represents a functional peptide  
 CC epitope of human heparanase.  
 XX  
 SQ Sequence 15 AA;  
 Query Match 100.0%; Score 82; DB 8; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SWELGNEPNSFLKKA 15  
 |||||  
 DB 1 SWELGNEPNSFLKKA 15  
 |||||

RESULT 3

ADU70829

XX ADU70829 standard; peptide; 15 AA.

AC ADU70829;

DT 10-FEB-2005 (first entry)

XX Human heparanase peptide SEQ ID NO:514.

DE enzyme; heparinase; vaccine; human leukocyte antigen; HLA;

KW immunostimulant; cytostatic; immune disorder; metastasis.

XX Homo sapiens.

PN EP1479764-A1.

PD 24-NOV-2004.

XX 19-MAY-2003; 2003EP-00011038.

XX 19-MAY-2003; 2003EP-00011038.

XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.

PA (UYHE-) UNIV HEIDELBERG RUPRECHT-KARLS.

XX Schirmacher V, Beckhove P, Sommerfeldt N;

XX WPI; 2005-014847/02.

XX New heparanase nonapeptide that binds to a human leukocyte antigen (HLA)

PT molecule or its functional derivative, useful for preparing a medicament

PT for inducing an immune response or for treating metastatic tumors.

PS Disclosure; SEQ ID NO 514; 269pp; English.

XX The invention relates to a novel heparanase peptide that binds to a human  
 CC leukocyte antigen (HLA) molecule, where the peptide is a nonapeptide, or  
 CC its functional derivative. A peptide of the invention has immunostimulant  
 CC and cytostatic activity, and is used in a vaccine. The heparinase peptide  
 CC is useful for preparing a medicament which induces an immune response or  
 CC for treating metastatic tumors. The present sequence represents a  
 CC heparinase peptide of the invention.  
 XX Sequence 15 AA;

Query Match 100.0%; Score 82; DB 9; Length 15;

Best Local Similarity 100.0%; Pred. No. 2.4e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15

DB 1 SWELGNEPNSFLKKA 15

|||||

RESULT 4

AEA42430

XX AEA42430 standard; peptide; 15 AA.

AC AEA42430;

DT 28-JUL-2005 (first entry)

XX Human heparanase epitope peptide SEQ ID NO:8.

DE antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;

KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;

KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;

KW angiogenesis disorder; cancer; tumor; metastasis; epitope.

XX Homo sapiens.

OS



XX AU2004201462-A1.  
 XX 06-MAY-2004.  
 XX 08-APR-2004; 2004AU-00201462.  
 XX 08-APR-2004; 2004AU-00201462.  
 XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.  
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.  
 XX Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;  
 PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;  
 PI Feinstein E;  
 XX WPI; 2005-173343/19.  
 XX Novel isolated antibody capable of specifically binding to epitope of  
 PT heparanase protein, useful for preventing and treating heparanase-related  
 PT disorder such as inflammatory disorder, scars, autoimmune conditions or  
 PT angiogenesis.  
 XX Claim 7; SEQ ID NO 8; 260pp; English.  
 XX The invention relates to an isolated antibody or its portion (I) capable  
 CC of specifically binding to an epitope of a heparanase protein. Also  
 CC described: (1) a cell line (II) for producing a monoclonal antibody or  
 CC its portion, comprising a cell line for producing (I); (2) a  
 CC pharmaceutical composition comprising (I) and a carrier; and (3) an  
 CC affinity medium (III) for binding human heparanase polypeptides,  
 CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)  
 CC useful for treating a subject suffering from a pathological condition,  
 CC which involves administering (I) to the subject. (I) is useful for  
 CC preventing and treating heparanase-related disorder or condition chosen  
 CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune  
 CC condition, angiogenesis, cell proliferation, cancerous condition, tumor  
 CC cell proliferation, invasion of circulating tumor cells and metastatic  
 CC disease. (I) is useful for detecting the presence of heparanase  
 CC polypeptide in a sample. (I) is useful for detecting heparanase-related  
 CC disease or condition in a subject such as vertebrate, preferably mammal  
 CC e.g., human. The heparanase-related disorder or condition further  
 CC includes renal disease or disorder chosen from diabetic nephropathy,  
 CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome  
 CC and renal cell carcinoma. The present sequence represents a human  
 CC heparanase epitope peptide, which is used in the exemplification of the  
 CC present invention.  
 XX Sequence 15 AA;  
 XX Query Match 100.0%; Score 82; DB 9; Length 15;  
 XX Best Local Similarity 100.0%; Pred. No. 2.4e-07;  
 XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SWELGNEPNFLKKA 15  
 DB 1 SWELGNEPNFLKKA 15  
 RESULT 5  
 ADR88207  
 ID ADR88207 standard; protein; 386 AA.  
 XX AC ADR88207;  
 XX DT 18-NOV-2004 (first entry)  
 XX DE Human mature heparanase 45 kDa major subunit.  
 XX Targeted drug delivery ; inflammatory disorder; wound; scar;  
 KW vasculopathy; autoimmune disorder; cancer; angiogenesis;  
 KW metastatic disease; atherosclerosis; restenosis; aneurysm; solid cancer;  
 KW non-solid cancer; haematopoietic malignancy ; lymphocytic leukaemia;

KW myelogenous leukaemia; Hodgkin's disease; multiple myeloma;  
 KW haemangiosarcoma; Kaposi's sarcoma; human ; heparanase; enzyme.  
 XX Homo sapiens.  
 XX US2004170631-A1.  
 XX 02-SEP-2004.  
 XX 28-NOV-2003; 2003US-00722502.  
 XX 02-SEP-1997; 97US-00922170.  
 PR 01-MAY-1998; 98US-00071739.  
 PR 04-NOV-1998; 98US-00186200.  
 PR 19-FEB-2003; 2003US-00368044.  
 PR 22-AUG-2003; 2003US-00645659.  
 XX (YACO/) YACOBY-ZEEVI O.  
 PA (PERE/) PERETZ T.  
 PA (MIRO/) MIRON D.  
 PA (SHLO/) SHLOMI Y.  
 PA (PECK/) PECKER I.  
 PA (AYAL/) AYAL-HERSHKOVITZ M.  
 PA (FEIN/) FEINSTEIN E.  
 PA (VGEL/) VAN GELDER J M.  
 PA (VLOD/) VLODAVSKY I.  
 PA (FRIE/) FRIEDMANN Y.  
 XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
 PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;  
 PI Friedmann Y;  
 XX WPI; 2004-625084/60.  
 XX Targeted drug delivery to a heparanase-expressing tissue of a patient,  
 PT useful for treating heparanase-associated conditions such as inflammation  
 PT or cancer, comprises administering a drug and an anti-heparanase antibody  
 PT complex.  
 XX Claim 2; SEQ ID NO 1; 58pp; English.  
 XX The invention relates to a method of targeted drug delivery to a tissue  
 CC of a patient, the tissue expressing heparanase. The method comprises  
 CC providing a complex of a drug directly or indirectly linked to an anti-  
 CC heparanase antibody, and administering the complex to the patient. In the  
 CC targeted drug delivery, the antibody comprises an antibody or its portion  
 CC capable of specifically binding to at least one epitope of a heparanase  
 CC protein. The composition and methods of the invention are useful for  
 CC diagnosing, preventing or treating conditions associated with heparanase  
 CC catalytic activity (e.g. an inflammatory disorder, wound, scar,  
 CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell  
 CC proliferation, invasion of circulating tumour cells and metastatic  
 CC disease), for purifying heparanase, or for developing drugs for those  
 CC heparanase-associated conditions. The vasculopathy is atherosclerosis,  
 CC restenosis or aneurysm. The cancerous condition is a solid cancer or a  
 CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy  
 CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous  
 CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous  
 CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,  
 CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and  
 CC multiple myeloma. The solid cancer is selected from tumours in lip and  
 CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,  
 CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,  
 CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of  
 CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue  
 CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,  
 CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic  
 CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary  
 CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,  
 CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,  
 CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,  
 CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's  
 CC sarcoma. The present sequence is the 45 kDa major subunit of human mature

CC heparanase.  
 XX Sequence 386 AA;  
 SQ

Query Match 100.0%; Score 82; DB 8; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 1e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SWELGNEPNSFLKKA 15  
 Db 62 SWELGNEPNSFLKKA 76

RESULT 6  
 ADT78174  
 ID ADT78174 standard; protein; 386 AA.  
 XX  
 AC  
 XX  
 DT 13-JAN-2005 (first entry)  
 XX  
 DE 45kDa subunit of mature processed human heparanase dimer.  
 XX  
 KW Antibody; epitope; heparanase; pathological condition; angiogenesis;  
 KW cell proliferation; cancerous condition; tumour cell invasion;  
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;  
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;  
 KW cystostatic; antiinflammatory; vulnery; antiarteriosclerotic;  
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2004213789-A1.  
 XX  
 PD 28-OCT-2004.  
 XX  
 PF 22-AUG-2003; 2003US-00645659.  
 XX  
 PR 02-SEP-1997; 97US-00922170.  
 PR 01-MAY-1998; 98US-00071739.  
 PR 04-NOV-1998; 98US-00186200.  
 PR 19-FEB-2003; 2003US-00368044.  
 XX  
 PA (YACO/) YACOBY-ZEEVI O.  
 PA (PERE/) PERETZ T.  
 PA (MIRO/) MIRON D.  
 PA (SHLO/) SHLOMI Y.  
 PA (PECK/) PECKER I.  
 PA (AYAL/) AYAL-HERSHKOVITZ M.  
 PA (FEIN/) FEINSTEIN E.  
 PA (GELD/) GELDER J M V.  
 PA (VLOD/) VLODAVSKY I.  
 PA (FRIE/) FRIEDMANN Y.  
 XX  
 PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
 PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;  
 PI Friedmann Y;  
 XX  
 DR WPI; 2004-774790/76.  
 XX  
 XX New neutralizing monoclonal anti-heparanase antibodies, useful for  
 PT detecting, treating or preventing cancer, inflammation or autoimmune  
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.  
 XX  
 PS Claim 5; SEQ ID NO 1; 68pp; English.  
 XX  
 CC The invention relates to an isolated antibody or antibody portion capable  
 CC of specifically binding to or elicited by at least one epitope of a  
 CC heparanase protein, where the heparanase protein is at least 60%  
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and  
 CC where at least one epitope comprises a sequence at least 70% homologous  
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)  
 CC a hybridoma cell line comprising a cell line for producing the monoclonal

antibody, (b) a method for detecting, treating or preventing a  
 pathological condition or a heparanase-related disorder or condition in a  
 subject, (c) a method for monitoring the state of a heparanase-related  
 disorder or condition in a subject, and (d) a pharmaceutical composition  
 comprising the isolated anti-heparanase antibody or antibody portion and  
 a pharmaceutical carrier. The antibody, methods, and composition are  
 useful for detecting, treating, preventing or monitoring a pathological  
 condition, e.g. angiogenesis, cell proliferation, a cancerous condition  
 (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,  
 or prostate cancer), minor cell proliferation, invasion of circulating  
 tumour cells, or a metastatic disease, or a heparanase-related disorder  
 or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy  
 (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or  
 renal disease or disorder (diabetic nephropathy, glomerulosclerosis,  
 nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell  
 carcinoma) in a mammal. This sequence represents the 45kDa subunit of  
 mature processed human heparanase dimer.  
 XX  
 SQ Sequence 386 AA;  
 Query Match 100.0%; Score 82; DB 8; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 1e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SWELGNEPNSFLKKA 15  
 Db 62 SWELGNEPNSFLKKA 76

RESULT 7  
 ADY27057  
 ID ADY27057 standard; protein; 386 AA.  
 XX  
 AC ADY27057;  
 XX  
 DT 05-MAY-2005 (first entry)  
 XX  
 DE Heparanase inhibitor protein #1.  
 XX  
 KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
 KW neurological disease; viral infection; infection; cytostatic;  
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
 KW heparanase modulator; enzyme purification.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2005016227-A2.  
 XX  
 PD 24-FEB-2005.  
 XX  
 PF 12-AUG-2004; 2004WO-IL000744.  
 XX  
 PR 14-AUG-2003; 2003US-0494800P.  
 PR 12-JAN-2004; 2004US-0535492P.  
 XX  
 PA (INST-) INSIGHT BIOPHARMACEUTICALS LTD.  
 XX  
 PI Van-Gelder JM, Miron D;  
 XX  
 DR WPI; 2005-182203/19.  
 XX  
 XX Regulating heparanase activity, useful for treating heparanase-associated  
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
 PT neurological diseases or viral diseases) comprises modulating heparanase  
 PT activation.  
 XX  
 PS Claim 55; SEQ ID NO 33; 211pp; English.  
 XX  
 CC The invention relates to a method of regulating heparanase activity in a  
 CC tissue or regulating a biological process depending at least in part on  
 CC heparanase activity comprising modulating heparanase activation. The  
 CC invention also relates to methods of treating a heparanase- or heparin  
 CC binding protein-associated disease or disorder in a subject, a

CC pharmaceutical composition for use in the treatment of a heparanase-associated disease or disorder comprising a therapeutic amount of an agent capable of modulating heparanase activation and a pharmaceutical carrier or diluent, a method of identifying a protease activator of heparanase, a protease substrate mimetic comprising a peptide representing a subset or all substrate residues or cleavage sites of human heparanase or an equivalent non-human heparanase, a method of producing active heparanase and a method of modulating an adhesion activity of heparanase. The composition and methods are useful for modulating heparanase activation and for treating heparanase-associated diseases or disorders such as cancer, inflammation, cardiovascular diseases, neurological diseases or viral infections. This sequence CC represents a heparanase inhibitor protein used in the scope of the invention.

XX

SQ Sequence 386 AA;

Query Match 100.0%; Score 82; DB 9; Length 386;  
Best Local Similarity 100.0%; Pred. No. 1e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
| | | | | | | | | | | | | | |  
Db 62 SWELGNEPNSFLKKA 76

RESULT 8

ID ADZ18995 standard; protein; 386 AA.

XX

AC ADZ18995;

XX

DT 16-JUN-2005 (first entry)

XX

DE Human heparanase consensus cleavage site #2.

XX

KW Enzyme engineering; heparanase; metastasis; autoimmune disease; inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic; immunosuppressive; enzyme.

KW

XX Homo sapiens.

OS

XX WO2005030962-A1.

PN

XX 07-APR-2005.

PD

XX

PF 17-SEP-2004; 2004WO-EP010517.

PP

XX

PR 26-SEP-2003; 2003US-0506479P.

PR

XX 20-JAN-2004; 2004US-0537729P.

XX

XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.

PA

XX Lahm A, Nardella C, Pallaro M, Steinkuhler C;

PI

XX WPI; 2005-273382/28.

DR

XX

XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a nucleotide sequence that encodes mammalian heparanase protein and has two consensus cleavage sites located between specific nucleotide encoding residues.

PT

PT

PT

XX

XX Disclosure; SEQ ID NO 16; 65pp; English.

PS

XX

XX The invention relates to a synthetic nucleic acid molecule that encodes mammalian heparanase protein, where the nucleic acid comprises two consensus cleavage sites recognized by endoproteinase. The sequences are useful for expressing mammalian heparanase in non-mammalian cells and in inhibitor screening assays for the development of therapeutics or pharmaceuticals for inhibiting or treating metastasis, autoimmune disease and/or inflammation. This sequence represents a human heparanase consensus cleavage site used in the scope of the invention.

CC

SQ Sequence 386 AA;

Query Match 100.0%; Score 82; DB 9; Length 386;  
Best Local Similarity 100.0%; Pred. No. 1e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
| | | | | | | | | | | | | | |  
Db 62 SWELGNEPNSFLKKA 76

RESULT 9

AEA42423

ID AEA42423 standard; protein; 386 AA.

XX

AC AEA42423;

XX

DT 28-JUL-2005 (first entry)

XX

DE Human mature heparanase dimer 45 kDa subunit SEQ ID NO:1.

XX

KW antibody; heparanase; antiinflammatory; vulnery; immunosuppressive; antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic; inflammation; wound healing; scarring; vasculopathy; autoimmune disease; angiogenesis disorder; cancer; tumor; metastasis.

KW

XX Homo sapiens.

OS

XX AU2004201462-A1.

PN

XX 06-MAY-2004.

PD

XX

PF 08-APR-2004; 2004AU-00201462.

PP

XX

PR 08-APR-2004; 2004AU-00201462.

PR

XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

PA

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX

PI Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;

PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;

PI Feinstein E;

XX

DR WPI; 2005-173343/19.

XX

XX Novel isolated antibody capable of specifically binding to epitope of heparanase protein, useful for preventing and treating heparanase-related disorder such as inflammatory disorder, scars, autoimmune conditions or angiogenesis.

PT

PT

PT

XX

PS Claim 2; SEQ ID NO 1; 260pp; English.

XX

XX The invention relates to an isolated antibody or its portion (I) capable of specifically binding to an epitope of a heparanase protein. Also described: (1) a cell line (II) for producing a monoclonal antibody or its portion, comprising a cell line for producing (I); (2) a pharmaceutical composition comprising (I) and a carrier; and (3) an affinity medium (III) for binding human heparanase polypeptides, comprising (I) immobilized to a chemically inert, insoluble carrier. (I) useful for treating a subject suffering from a pathological condition, which involves administering (I) to the subject. (I) is useful for preventing and treating heparanase-related disorder or condition chosen from inflammatory disorder, wound, scar, vasculopathy, autoimmune condition, angiogenesis, cell proliferation, cancerous condition, tumor cell proliferation, invasion of circulating tumor cells and metastatic disease. (I) is useful for detecting the presence of heparanase polypeptide in a sample. (I) is useful for detecting heparanase-related disease or condition in a subject such as vertebrate, preferably mammal e.g., human. The heparanase-related disorder or condition further includes renal disease or disorder chosen from diabetic nephropathy, glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome and renal cell carcinoma. The present sequence represents the 45 kDa subunit of the human mature processed heparanase dimer, which is used in

SQ Sequence 486 AA;

SQ Sequence 486 AA;

Query Match 100.0%; Score 82; DB 9; Length 486;  
Best Local Similarity 100.0%; Pred. No. 1.3e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
| | | | | | | | | | |  
Db 184 SWELGNEPNSFLKKA 198

RESULT 12  
ADZ18996  
ID ADZ18996 standard; protein; 492 AA.

AC ADZ18996;  
XX  
XX  
XX 16-JUN-2005 (first entry)  
XX  
XX Hep106 construct protein.  
XX  
XX Enzyme engineering; heparanase; hep106; metastasis; autoimmune disease;  
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
KW immunosuppressive; enzyme.  
XX  
XX Synthetic.  
XX WO2005030962-A1.  
XX  
XX  
XX 07-APR-2005.  
XX  
XX 17-SEP-2004; 2004WO-EP010517.  
XX  
XX 26-SEP-2003; 2003US-0506479P.  
XX 20-JAN-2004; 2004US-0537729P.  
XX

PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.  
XX  
XX Lahm A, Nardella C, Pallaoaro M, Steinkuhler C;  
XX  
XX WPI; 2005-273382/28.  
DR N-PSDB; ADZ18997.  
XX  
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
PT nucleotide sequence that encodes mammalian heparanase protein and has two  
PT consensus cleavage sites located between specific nucleotide encoding  
PT residues.  
XX  
XX Example 2; SEQ ID NO 17; 65pp; English.

PS  
XX The invention relates to a synthetic nucleic acid molecule that encodes  
XX mammalian heparanase protein, where the nucleic acid comprises two  
CC consensus cleavage sites recognized by endoproteinase. The sequences are  
CC useful for expressing mammalian heparanase in non-mammalian cells and in  
CC inhibitor screening assays for the development of therapeutics or  
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
CC and/or inflammation. This sequence represents a hep106 construct protein  
CC used in the scope of the invention.  
XX  
XX

SQ Sequence 492 AA;

Query Match 100.0%; Score 82; DB 9; Length 492;  
Best Local Similarity 100.0%; Pred. No. 1.3e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
| | | | | | | | | | |  
Db 168 SWELGNEPNSFLKKA 182

RESULT 13  
AEB87562  
ID AEB87562 standard; protein; 493 AA.  
XX

AC AEB87562;  
XX  
XX 06-OCT-2005 (first entry)  
XX  
XX Human heparanase 65delta15 deletion mutant.  
XX  
XX Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;  
KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;  
KW leukemia; lymphoma; dermatological disease; hematological disease;  
KW immune disorder; inflammation; antiinflammatory; renal disease;  
KW nephrotropic; endocrine disease; genitourinary disease;  
KW autoimmune disease; immunosuppressive; drug screening; mutein.  
XX  
XX Homo sapiens.  
OS Synthetic.  
OS  
XX WO2005071070-A2.  
PN  
XX 04-AUG-2005.  
XX  
XX 20-JAN-2005; 2005WO-IL000068.  
XX  
XX 22-JAN-2004; 2004IL-00160025.  
PR 28-JUL-2004; 2004US-00901943.  
PR  
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.  
XX  
XX Vlodavsky I, Ilan N, Levy-Adam F;  
PI  
XX WPI; 2005-564219/57.  
DR N-PSDB; AEB87561.  
DR  
XX New amino acid sequences derived from the 50 kDa subunit of heparanase,  
PT for treating or inhibiting malignant proliferative, inflammatory, kidney  
PT disorder or autoimmune disorder.  
XX  
XX Claim 105; SEQ ID NO 4; 167pp; English.

PS  
XX The present sequence is that of a deletion mutant of human heparanase,  
XX denoted 65delta15, which is devoid of amino acid residues 158-171 of the  
CC native protein. The recombinant protein is deficient of heparanase  
CC endoglycosidase catalytic activity. The invention relates to amino acid  
CC sequences derived from the N-terminus region of the 50 kDa subunit of  
CC heparanase, particularly in the regions between amino acid residues 158-  
CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise  
CC heparin-binding domains. The invention also provides an antibody directed  
CC to these sequences, in particular the 158-171 peptide, and compositions  
CC and uses of this antibody as a heparanase inhibitor. A screening method  
CC is provided for specific heparanase inhibitors. Claimed pharmaceutical  
CC compositions comprising (i) a peptide derived from the N-terminus region  
CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a  
CC peptide, or (iii) an antibody which specifically recognizes the peptide  
CC are used for the inhibition of heparanase catalytic activity associated  
CC with an inflammatory disorder, kidney disease, autoimmune disease,  
CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or  
CC with a malignant proliferative disorder, especially a solid or non-solid  
CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.  
XX  
XX

SQ Sequence 493 AA;

Query Match 100.0%; Score 82; DB 9; Length 493;  
Best Local Similarity 100.0%; Pred. No. 1.3e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
| | | | | | | | | | |  
Db 169 SWELGNEPNSFLKKA 183

RESULT 14  
ADZ18999  
ID ADZ18999 standard; protein; 495 AA.  
XX

AC AD218999;  
 XX  
 DT 16-JUN-2005 (first entry)  
 XX  
 DE Hep109 construct protein.  
 XX  
 KW Enzyme engineering; heparanase; hepl109; metastasis; autoimmune disease;  
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
 KW immunosuppressive; enzyme.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2005030962-A1.  
 XX  
 PD 07-APR-2005.  
 XX  
 PF 17-SEP-2004; 2004WO-EP010517.  
 XX  
 PR 26-SEP-2003; 2003US-0506479P.  
 PR 20-JAN-2004; 2004US-0537729P.  
 XX  
 PA (RICE-) 1ST RICERCHE BIOL MOLECOLARE ANGELETTI.  
 XX  
 PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;  
 XX  
 DR WPI; 2005-273382/28.  
 DR N-PSDB; AD218998.  
 XX  
 PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.  
 XX  
 PS Example 2; SEQ ID NO 20; 65pp; English.  
 XX  
 CC The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoproteinase. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in  
 CC inhibitor screening assays for the development of therapeutics or  
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a hepl109 construct protein  
 CC used in the scope of the invention.  
 XX  
 SQ Sequence 495 AA;  
 Query Match 100.0%; Score 82; DB 9; Length 495;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-05; Mismatches 0; Indels 0; Gaps 0;  
 Matches 15; Conservative 0;  
 QY 1 SWELGNPNFLKKA 15  
 DB 171 SWELGNPNFLKKA 185  
 |||||  
 RESULT 15  
 AEB87587  
 ID AEB87587 standard; protein; 497 AA.  
 XX  
 AC AEB87587;  
 XX  
 DT 06-OCT-2005 (first entry)  
 XX  
 DE Human heparanase 65delta10 deletion mutant.  
 XX  
 KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;  
 KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;  
 KW leukemia; lymphoma; dermatological disease; hematological disease;  
 KW immune disorder; inflammation; antiinflammatory; renal disease;  
 KW nephrotropic; endocrine disease; genitourinary disease;  
 KW autoimmune disease; immunosuppressive; drug screening; mutcin.  
 XX  
 OS Homo sapiens.  
 XX

OS Synthetic.  
 XX  
 PN WO2005071070-A2.  
 XX  
 PD 04-AUG-2005.  
 XX  
 PF 20-JAN-2005; 2005WO-IL000068.  
 XX  
 PR 22-JAN-2004; 2004IL-00160025.  
 PR 28-JUL-2004; 2004US-00901943.  
 XX  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.  
 XX  
 PI Vlodavsky I, Ilan N, Levy-Adam F;  
 XX  
 DR WPI; 2005-564219/57.  
 DR N-PSDB; AEB87586.  
 XX  
 PT New amino acid sequences derived from the 50 kDa subunit of heparanase,  
 PT for treating or inhibiting malignant proliferative, inflammatory, kidney  
 PT disorder or autoimmune disorder.  
 XX  
 PS Claim 106; SEQ ID NO 29; 167pp; English.  
 XX  
 CC The present sequence is that of a deletion mutant of human heparanase,  
 CC denoted 65delta10, which is devoid of amino acid residues 270-280 of the  
 CC native protein. The recombinant protein is deficient of heparanase  
 CC endoglycosidase catalytic activity. The invention relates to amino acid  
 CC sequences derived from the N-terminus region of the 50 kDa subunit of  
 CC heparanase, particularly in the regions between amino acid residues 158-  
 CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise  
 CC heparin-binding domains. The invention also provides an antibody directed  
 CC to these sequences, in particular the 158-171 peptide, and compositions  
 CC and uses of this antibody as a heparanase inhibitor. A screening method  
 CC is provided for specific heparanase inhibitors. Claimed pharmaceutical  
 CC compositions comprising (i) a peptide derived from the N-terminus region  
 CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a  
 CC peptide, or (iii) an antibody which specifically recognizes the peptide  
 CC are used for the inhibition of heparanase catalytic activity associated  
 CC with an inflammatory disorder, kidney disease, autoimmune disease,  
 CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or  
 CC with a malignant proliferative disorder, especially a solid or non-solid  
 CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.  
 XX  
 SQ Sequence 497 AA;  
 Query Match 100.0%; Score 82; DB 9; Length 497;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-05; Mismatches 0; Indels 0; Gaps 0;  
 Matches 15; Conservative 0;  
 QY 1 SWELGNPNFLKKA 15  
 DB 184 SWELGNPNFLKKA 198  
 |||||  
 RESULT 16  
 AD219000  
 ID AD219000 standard; protein; 501 AA.  
 XX  
 AC AD219000;  
 XX  
 DT 16-JUN-2005 (first entry)  
 XX  
 DE HepGS3 construct protein.  
 XX  
 KW Enzyme engineering; heparanase; hepGS3; metastasis; autoimmune disease;  
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
 KW immunosuppressive; enzyme.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2005030962-A1.  
 XX





CC agent capable of modulating heparanase activation and a pharmaceutical  
 CC carrier or diluent, a method of identifying a protease activator of  
 CC heparanase, a protease substrate mimetic comprising a peptide  
 CC representing a subset or all substrate residues or cleavage sites of  
 CC human heparanase or an equivalent non-human heparanase, a method of  
 CC producing active heparanase and a method of modulating an adhesion  
 CC activity of heparanase. The composition and methods are useful for  
 CC modulating heparanase activation and for treating heparanase-associated  
 CC diseases or disorders such as cancer, inflammation, cardiovascular  
 CC diseases, neurological diseases or viral infections. This sequence  
 CC represents a human inactive heparanase protein used in the scope of the  
 CC invention.

XX SQ Sequence 508 AA;

Query Match 100.0%; Score 82; DB 9; Length 508;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
 |||||

Db 184 SWELGNEPNSFLKKA 198

RESULT 19

ID ADZ19006 standard; protein; 526 AA.

XX AC ADZ19006;

XX DT 16-JUN-2005 (first entry)

XX DE HepHyaluro construct protein.

XX KW Enzyme engineering; heparanase; hepHyaluro; metastasis;  
 KW autoimmune disease; inflammation; neoplasm; immune disorder;  
 KW antiinflammatory; cytostatic; immunosuppressive; enzyme.

XX OS Synthetic.

XX PN WO2005030962-A1.

XX PD 07-APR-2005.

XX PF 17-SEP-2004; 2004WO-EP010517.

XX PR 26-SEP-2003; 2003US-0506479P.

XX PR 20-JAN-2004; 2004US-0537729P.

XX PA (RICE-) IST RICERCH BIOL MOLECOLARE ANGELETTI.

XX PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;

XX WPI; 2005-273382/28.

XX DR N-PSDB; ADZ19007.

XX PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.

XX PS Example 2; SEQ ID NO 27; 65pp; English.

XX CC The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoproteinase. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in  
 CC inhibitor screening assays for the development of therapeutic or  
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a hepHyaluro construct  
 CC protein used in the scope of the invention.

XX SQ Sequence 526 AA;

Query Match 100.0%; Score 82; DB 9; Length 526;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
 |||||

Db 202 SWELGNEPNSFLKKA 216

RESULT 20

ID ABB07815 standard; protein; 527 AA.

XX AC ABB07815;

XX DT 03-JUL-2002 (first entry)

XX DE Chicken signal peptide/human heparanase chimeric protein sequence.

XX KW Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;  
 KW anti-protozoan; neuroprotective; heparin; chicken; human; chimeric.

XX OS Synthetic.

XX OS Gallus gallus.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..19

FT Protein /note= "chicken heparanase signal peptide"

FT /note= "human heparanase mature protein"

XX PN US2002034810-A1.

XX PD 21-MAR-2002.

XX PF 16-AUG-2001; 2001US-00930218.

XX PR 20-SEP-2000; 2000US-00666390.

XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX PI Goldshmidt O, Pecker I, Vlodaysky I, Michal I, Zcharia E;

XX WPI; 2002-338926/37.

XX DR N-PSDB; ABL40753.

XX PT Nucleic acid encoding avian and reptile heparanase polypeptide is useful  
 PT to treat various heparin-related disorders and the signal peptide is  
 PT useful in production of membrane-targeted or secreted recombinant  
 PT proteins.

XX PS Disclosure; Page 26-28; 39pp; English.

XX CC The invention relates to an isolated avian and reptile nucleic acid,  
 CC encoding a polypeptide with heparanase catalytic activity. The signal  
 CC peptide of the nucleic acid can be used to express membrane-associated or  
 CC secreted proteins in heterologous expression systems. The encoded  
 CC polypeptides can be used to prevent tumour angiogenesis, metastasis and  
 CC invasion, and to intervene with pathologies associated with impaired  
 CC heparin-binding growth factors, cellular responses to heparin-binding  
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,  
 CC cellular susceptibility to viral, protozoan and bacterial infections or  
 CC disintegration of neurodegenerative plaques. The present sequence  
 CC represents a chicken signal peptide/human heparanase chimeric protein  
 CC sequence

XX SQ Sequence 527 AA;

Query Match 100.0%; Score 82; DB 5; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
Db 203 SWELGNEPNSFLKKA 217

RESULT 21  
ABW02018  
ID ABW02018 standard; protein; 527 AA.

XX AC ABW02018;  
XX AC ABW02018;

DT 12-FEB-2004 (first entry)

XX Chimeric human-chicken heparanase protein.

KW Chicken; heparanase; tumour cell metastasis; inflammation; autoimmunity;  
KW wound healing; angiogenesis; restenosis; Genstmann-Straussler Syndrome;  
KW neurodegenerative disease; atherosclerosis; Creutzfeldt-Jakob disease;  
KW infection; Scrapie; Alzheimer's disease; protein therapy; cytostatic;  
KW immunosuppressive; vulnerary; bactericide; anti-angiogenic; virucide;  
KW antisclerotic; neuroprotective; protozoacide; chimeric; fusion protein;  
KW enzyme; human.

XX Chimeric - Gallus gallus.  
OS Chimeric - Homo sapiens.

XX US2003180788-A1.

XX 25-SEP-2003.

XX 08-MAY-2003; 2003US-00431438.

XX 20-SEP-2000; 2000US-00666390.

PR 16-AUG-2001; 2001US-00930218.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

PA Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;

XX WPI; 2003-843931/78.

XX N-PSDB; AAD63532.

XX Recombinant jungle red fowl (Gallus gallus) heparanase protein, useful  
PT for treating cancers, microbial infections and aiding wound healing.

XX Example; Page 26-28; Opp; English.

XX The present invention relates to novel jungle red fowl heparanase protein  
CC and polynucleotides encoding such proteins. Heparanase sequences can be  
CC used to develop treatments for various diseases, to develop diagnostic  
CC assays for these diseases and to provide new tools for basic and directed  
CC research especially in the fields of medicine and biology. They can be  
CC used to develop new drugs to inhibit tumour cell metastasis, inflammation  
CC and autoimmunity. Recombinant heparanase offers a potential treatment for  
CC wound healing, angiogenesis, restenosis, atherosclerosis, inflammation,  
CC neurodegenerative diseases (e.g. Genstmann-Straussler Syndrome, Scrapie,  
CC Creutzfeldt-Jakob disease and Alzheimer's disease) and certain viral and  
CC some bacterial and protozoa infections. Recombinant heparanase can also  
CC be used to neutralise plasma heparin, as a potential replacement of  
CC protamine. Sequences of the invention are used in protein therapy. The  
CC present sequence is chimeric human-chicken heparanase protein

XX Sequence 527 AA;

Query Match 100.0%; Score 82; DB 7; Length 527;

Best Local Similarity 100.0%; Pred. No. 1.4e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
Db 203 SWELGNEPNSFLKKA 217

CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialisation of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be use in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte  
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents a chimeric  
 CC protein comprising the signal peptide of chicken heparanase and residues  
 CC 35-543 of the human heparanase mutant E343A.

XX SQ Sequence 527 AA;

Query Match 100.0%; Score 82; DB 8; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SWELGNEPNSFLKKA 15  
 |||||  
 Db 203 SWELGNEPNSFLKKA 217

RESULT 23

AD219004  
 ID AD219004 standard; protein; 527 AA.

XX AC AD219004;

XX DT 16-JUN-2005 (first entry)

XX DE HepGS4 construct protein.

XX KW Enzyme engineering; heparanase; hepGS4; metastasis; autoimmune disease;  
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
 KW immunosuppressive; enzyme.

XX OS Synthetic.

XX PN WO2005030962-A1.

XX PD 07-APR-2005.

XX PF 17-SEP-2004; 2004WO-EP010517.

XX PR 26-SEP-2003; 2003US-0506479P.

XX PR 20-JAN-2004; 2004US-0537729P.

XX PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.

XX PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;

XX DR WPI; 2005-273382/28.

XX DR N-PSDB; AD219002.

XX PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.

XX PS Example 2; SEQ ID NO 25; 65pp; English.

XX CC The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoproteinase. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in  
 CC inhibitor screening assays for the development of therapeutics or  
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a hepGS4 construct protein  
 CC used in the scope of the invention.

SQ Sequence 527 AA;

Query Match 100.0%; Score 82; DB 9; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SWELGNEPNSFLKKA 15  
 |||||  
 Db 203 SWELGNEPNSFLKKA 217

RESULT 24

AAV34173

ID AAY34173 standard; protein; 530 AA.

XX AC AAY34173;

XX DT 15-NOV-1999 (first entry)

XX DE Human pre-proheparanase protein sequence.

XX KW Human; pre-proheparanase; platelet; wound healing; angiogenesis blocker;  
 KW inflammation; psoriasis; diabetic retinopathy; solid tumour; arthritis;  
 KW heparin degradation; anticoagulant neutralisation; asthma; CNS disease;  
 KW inflammatory disease; vascular restenosis; atherosclerosis; diagnosis;  
 KW tumour growth; fibroproliferative disorder; neurodegenerative disease;  
 KW therapy.

XX OS Homo sapiens.

XX PN WO9943830-A2.

XX PD 02-SEP-1999.

XX PF 18-FEB-1999; 99WO-US001489.

XX PR 24-FEB-1998; 98US-0075706P.

XX PR 26-MAR-1998; 98US-0079401P.

XX PA (PHAA ) PHARMACIA & UPJOHN CO.

XX PI Heirnikson RL, Fairbanks MB, Mildner AM;

XX DR WPI; 1999-540598/45.

XX DR N-PSDB; AAZ11236.

XX PT New isolated platelet heparanase polypeptides, used to develop products  
 for, e.g. wound healing and blocking angiogenesis.

XX PS Claim 12; Fig 7; 57pp; English.

XX CC This sequence is the human pre-proheparanase of the invention. This  
 CC sequence was isolated from human platelets. The heparanase can be used  
 CC for identifying agents which alter heparanase activity. The heparanase  
 CC can be used for wound healing or for blocking angiogenesis or  
 CC inflammation. It can be used for treating e.g. psoriasis, diabetic  
 CC retinopathy or solid tumours, or for the degradation of heparin and the  
 CC neutralisation of heparin's anticoagulant properties during surgery.  
 CC Inhibitors of heparanase activity can be used in the treatment of  
 CC atherosclerosis, asthma, and other inflammatory diseases, vascular restenosis,  
 CC disorders, and central nervous system (CNS) and neurodegenerative  
 CC diseases. The products can also be used for detection and diagnosis. The  
 CC purified heparanase, both recombinantly produced human heparanase and  
 CC heparanase isolated from human platelet activity, allows for the  
 CC convenient selection of compounds having anti-heparanase activity, i.e.  
 CC inhibitors of heparanase activity, by measuring inhibition of heparanase  
 CC activity. Inhibition of heparanase activity can be measured by blocking  
 CC heparanase-mediated release of radioactive fragments from in vivo  
 CC radiolabelled (HSPG)/heparin

XX SQ Sequence 530 AA;

Query Match 100.0%; Score 82; DB 2; Length 530;  
Best Local Similarity 100.0%; Pred. No. 1.4e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
|||  
Db 206 SWELGNEPNSFLKKA 220

## RESULT 25

AAAY17083  
ID AAY17083 standard; protein; 532 AA.

XX AC AAY17083;  
XX DT 21-JUL-1999 (first entry)  
XX DE  
XX OS Homo sapiens.  
XX PN WO9921975-A1.  
XX KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;  
XX KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;  
XX KW arteriosclerosis; atherosclerosis; inflammation; tissue development;  
XX KW human; HSPG.

XX OS Homo sapiens.

XX PN WO9921975-A1.

XX PD 06-MAY-1999.

XX PF 28-OCT-1998; 98WO-AU000898.

XX PR 28-OCT-1997; 97AU-00000062.

XX PR 09-DEC-1997; 97AU-00000812.

XX PA (AUSU ) UNIV AUSTRALIAN NAT.

XX PI Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;

XX DR WPI; 1999-312956/26.  
XX DR N-PSDB; AAX37260.

XX PT Polynucleotides encoding mammalian endoglucuronidases, especially  
XX PT heparanases, useful to promote wound healing.

XX PS Claim 6; Page 76-79; 112pp; English.

XX CC The invention relates to nucleic acid sequences that encode heparanase  
XX CC enzymes having endoglucuronidase activity. Recombinant heparanases are  
XX CC capable of removing the HS side chain from heparan sulfate proteoglycan  
XX CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to  
XX CC inhibit heparanase, this is useful for treatment of a physiological or  
XX CC medical condition associated with elevated heparanase activity, such as  
XX CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,  
XX CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and  
XX CC rat heparanases can be used to enhance wound healing, especially  
XX CC associated with tissue development and repair. The conditions mentioned  
XX CC above can be diagnosed using specific antibodies, and also using primers  
XX CC and probes specific for the heparanase polynucleotides. Other uses of the  
XX CC heparanases include sequencing sulfated molecules such as HSPG

XX SQ Sequence 532 AA;

Query Match 100.0%; Score 82; DB 2; Length 532;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
|||  
Db 219 SWELGNEPNSFLKKA 233

## RESULT 26

AAAY02345  
ID AAY02345 standard; protein; 543 AA.

XX AC AAY02345;

XX DT 09-JUL-1999 (first entry)

XX DE A human heparanase protein.

XX KW Heparanase; hp; modulator; heparin-binding growth factor;  
XX KW cellular response; cytokine; cell interaction; plasma lipoprotein;  
XX KW cellular susceptibility; infection; disintegration;  
XX KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;  
XX KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;  
XX KW plasma heparin; micrometastasis; autoimmune lesion; renal failure.

XX OS Homo sapiens.

XX PN WO9911798-A1.

XX PD 11-MAR-1999.

XX PF 31-AUG-1998; 98WO-US017954.

XX PR 02-SEP-1997; 97US-00922170.

XX PR 02-JUL-1998; 98US-00109386.

XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX PI Pecker I, Vlodaysky I, Feinstein E;

XX DR WPI; 1999-302255/25.  
XX DR N-PSDB; AAX35648.

XX PT New human polynucleotide useful for treating angiogenesis, restenosis,  
XX PT and inflammation.

XX PS Claim 6; Fig 1; 63pp; English.

XX CC The specification describes a polypeptide having heparanase (hp)  
XX CC activity. The recombinant protein is used as a modulator of heparin-  
XX CC binding growth factors, cellular responses to heparin-binding growth  
XX CC factors and cytokines, cell interaction with plasma lipoproteins  
XX CC cellular susceptibility to viral, protozoal and bacterial infections or  
XX CC disintegration of neurodegenerative plaques. Heparanase may be useful for  
XX CC conditions such as wound healing, angiogenesis, restenosis,  
XX CC atherosclerosis, inflammation, neurodegenerative diseases, and viral  
XX CC infections. Mammalian heparanase can be used to neutralize plasma  
XX CC heparin, and anti-heparanase antibodies may be applied for  
XX CC immunodetection and diagnosis of micrometastases, autoimmune lesions, and  
XX CC renal failure in biopsy specimens, plasma samples, and body fluids. The  
XX CC present sequence represents human heparanase

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 2; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
|||  
Db 219 SWELGNEPNSFLKKA 233

## RESULT 27

AAAY17082  
ID AAY17082 standard; protein; 543 AA.

XX AC AAY17082;

XX DT 21-JUL-1999 (first entry)

```

XX DE Human heparanase enzyme.
XX OS
XX KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
XX KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;
XX KW arteriosclerosis; atherosclerosis; inflammation; tissue development;
XX KW human; HSPG.
XX OS
XX OS Homo sapiens.
XX PN WO9921975-A1.
XX PR 06-MAY-1999.
XX PF 28-OCT-1998; 98WO-AU000898.
XX PR 28-OCT-1997; 97AU-00000062.
XX PR 09-DEC-1997; 97AU-00000812.
XX PA (AUSU ) UNIV AUSTRALIAN NAT.
XX PI Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;
XX DR WPI; 1999-312956/26.
XX DR N-PSDB; AAX37259.
XX PT Polynucleotides encoding mammalian endoglucuronidases, especially
XX PT heparanases, useful to promote wound healing.
XX PS Claim 6; Page 69-73; 112pp; English.
XX CC The invention relates to nucleic acid sequences that encode heparanase
XX CC enzymes having endoglucuronidase activity. Recombinant heparanases are
XX CC capable of removing the HS side chain from heparan sulfate proteoglycan
XX CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to
XX CC inhibit heparanase, this is useful for treatment of a physiological or
XX CC medical condition associated with elevated heparanase activity, such as
XX CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,
XX CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and
XX CC rat heparanases can be used to enhance wound healing, especially
XX CC associated with tissue development and repair. The conditions mentioned
XX CC above can be diagnosed using specific antibodies, and also using primers
XX CC and probes specific for the heparanase polynucleotides. Other uses of the
XX CC heparanases include sequencing sulfated molecules such as HSPG. The
XX CC present sequence represents a human heparanase
XX SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 2; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
DB 219 SWELGNEPNSFLKKA 233

RESULT 28
AAY57590
ID AAY57590 standard; protein; 543 AA.
XX
AC AAY57590;
XX
DT 02-MAR-2000 (first entry)
XX
DE Human heparanase.
XX
KW Human; heparanase; hpa; genetic modification; expression; anticancer;
KW angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumour;
KW anti-atherosclerotic; anti-inflammatory; antineurodegeneration;
KW heparan sulphate; heparin-binding growth factor; tumour angiogenesis;
KW metastasis; wound healing; restenosis; atherosclerosis; inflammation;
KW neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;
XX

```

---

```

KW micrometastasis; autoimmune lesion; kidney failure.
XX
OS Homo sapiens.
XX PN WO9957244-A1.
XX PD 11-NOV-1999.
XX PF 29-APR-1999; 99WO-US009256.
XX PR 01-MAY-1998; 98US-00071618.
XX PR 02-MAR-1999; 99US-00260038.
XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX PA (FRIE/) FRIEDMAN M M.
XX PI Ben-Artzi H, Ayal-Hershkovitz M, Yacoby-Zeevi O, Pecker I;
XX PI Peleg Y, Shlomi Y;
XX DR WPI; 2000-062144/05.
XX DR N-PSDB; AAZ39195.
XX PT Engineered cells that express recombinant heparanase, useful
XX PT therapeutically, e.g. for treating angiogenesis and to screen for
XX PT specific inhibitors, potential anticancer agents.
XX PS Claim 3; Page 107-109; 118pp; English.
XX CC The present invention describes genetically modified cells (A) containing
XX CC a polynucleotide (I) that encodes a polypeptide with heparanase activity,
XX CC and express recombinant heparanase (II). Heparanase cleaves heparan
XX CC sulphate (HS) at specific intrachain sites, resulting in release of
XX CC heparin-binding growth factors, enzymes and proteins that are sequestered
XX CC by HS in basement membranes, extracellular matrix or cell surfaces. It
XX CC may also be implicated in tumour angiogenesis and metastases. (II) is
XX CC potentially useful in wound healing and for treating angiogenesis,
XX CC restenosis, atherosclerosis, inflammation, neurodegeneration, viral
XX CC infection and cystic fibrosis. It can also be used to neutralise heparin
XX CC (an alternative to protamine) and to screen for specific inhibitors
XX CC (potentially useful for treating cancer and metastases). Antibodies
XX CC raised against (II) are used for immunodetection and diagnosis of
XX CC micrometastases, autoimmune lesions and kidney failure. (A) provide (II)
XX CC in large quantities, in a form that is homogeneously processed and
XX CC activated/neutralised by a dedicated protease. The present sequence
XX CC represents human heparanase
XX SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 3; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
DB 219 SWELGNEPNSFLKKA 233

RESULT 29
AAB08849
ID AAB08849 standard; protein; 543 AA.
XX
AC AAB08849;
XX
DT 15-JAN-2001 (first entry)
XX
DE Amino acid sequence of a human heparanase polypeptide.
XX
KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Strausler Syndrome; Creutzfeldt-Jakob disease.
XX

```

OS Homo sapiens.  
 XX WO200052178-A1.  
 PN  
 XX  
 XX 08-SEP-2000.  
 PD  
 XX  
 PF 14-FEB-2000; 2000WO-US003542.  
 XX  
 XX 01-MAR-1999; 99US-00258892.  
 PR  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA (FRIE/) FRIEDMAN M M.  
 XX  
 XX Pecker I, Vlodavsky I, Feinstein E;  
 PI  
 XX WPI; 2000-579289/54.  
 DR N-PSDB; AAA75051.  
 XX  
 XX New polynucleotides encoding a polypeptide having heparanase activity,  
 PT useful in wound healing and in gene therapy, particularly in treating  
 PT tumor, inflammation, autoimmunity, neurodegenerative diseases.  
 XX  
 PS Claim 22; Fig 1; 152pp; English.  
 XX  
 XX The present sequence represents a human protein with heparanase catalytic  
 CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,  
 CC particularly in treating tumour, inflammation or autoimmunity.  
 CC Particularly, the polynucleotide is useful in modulating the  
 CC bioavailability of heparin-binding growth factors, cellular responses to  
 CC heparin-binding growth factors (e.g. bFGF) and cytokines (e.g.  
 CC interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular  
 CC susceptibility to certain viral and some bacterial and protozoa  
 CC infections, or disintegration of neurodegenerative plaques. The  
 CC polynucleotide is also useful in wound healing (e.g. thermal, chemical or  
 CC radiation burns), and in the treatment of angiogenesis, restenosis,  
 CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-  
 CC Strausler Syndrome or Creutzfeldt-Jakob disease), and some viral,  
 CC bacterial or protozoa infections  
 XX  
 SQ Sequence 543 AA;  
 Query Match 100.0%; Score 82; DB 3; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SWELGNEPNSFLKKA 15  
 |||||  
 DB 219 SWELGNEPNSFLKKA 233  
 RESULT 30  
 AAY52990  
 ID AAY52990 standard; protein; 543 AA.  
 XX  
 AC AAY52990;  
 XX  
 XX 21-FEB-2000 (first entry)  
 DT  
 XX Human heparanase protein sequence.  
 DE  
 XX Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;  
 KW antidiabetic; immunomodulatory; anti-inflammation; nephrotropic;  
 KW metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;  
 KW mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;  
 KW inflammation; haemorrhagic nephritis; nephrotic syndrome;  
 KW autoimmune disease; anticancer; kidney disease.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO9957153-A1.  
 PN  
 XX 11-NOV-1999.  
 PD

XX 29-APR-1999; 99WO-US009255.  
 PF  
 XX 01-MAY-1998; 98US-00071739.  
 PR  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA (FRIE/) FRIEDMAN M M.  
 XX  
 XX Pecker I, Vlodavsky I, Friedman Y, Perets T;  
 PI  
 XX WPI; 2000-052944/04.  
 XX N-PSDB; AAZ33290.  
 DR  
 XX Heparanase-specific molecular probes useful for diagnosis and treatment,  
 PT e.g. of tumors, and for targeted drug delivery.  
 PT  
 XX Example; Page 81-82; 90pp; English.  
 PS  
 XX The present invention describes heparanase-specific molecular probes,  
 CC useful for methods of detecting heparanase in situ. The probes and anti-  
 CC heparanase antibodies are used to detect or quantify the expression of  
 CC heparanase, for diagnosis and monitoring of diseases (especially  
 CC metastasis), for treatment of heparanase-associated diseases (e.g.  
 CC tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,  
 CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its  
 CC metastases) derived from liver, prostate, bladder, breast, ovary, cervix,  
 CC colon, skin, intestine, stomach, uterus and pancreas, kidney disease,  
 CC diabetes and inflammation, haemorrhagic nephritis, nephrotic syndrome,  
 CC sepsis and inflammatory or autoimmune disease), for targeted drug  
 CC delivery (e.g. of anticancer agents) and as research reagents. The  
 CC present sequence represents human heparanase, which is used in the  
 CC exemplification of the present invention  
 XX  
 SQ Sequence 543 AA;  
 Query Match 100.0%; Score 82; DB 3; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SWELGNEPNSFLKKA 15  
 |||||  
 DB 219 SWELGNEPNSFLKKA 233  
 RESULT 31  
 AAY97635  
 ID AAY97635 standard; protein; 543 AA.  
 XX  
 AC AAY97635;  
 XX  
 XX 20-APR-2001 (first entry)  
 DT  
 XX Human heparanase protein sequence.  
 DE  
 XX Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scarpe;  
 KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;  
 KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;  
 KW gene therapy; human.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200100643-A2.  
 PN  
 XX 04-JAN-2001.  
 PD  
 XX 19-JUN-2000; 2000WO-IL000358.  
 PF  
 XX 25-JUN-1999; 99US-0140801P.  
 PR  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA  
 XX Pecker I, Michal I, Itzhaki H;  
 PI

XX WPI; 2001-137930/14.  
 XX New polynucleotides and polypeptides that are distantly homologous to  
 PT heparanase, useful in wound healing, as well as in gene therapy protocols  
 PT for angiogenesis, restenosis, atherosclerosis, or inflammation.  
 XX  
 PS Disclosure; Page 64-65; 67pp; English.  
 XX  
 PS This sequence represents a heparanase of the invention. The heparanase  
 CC DNA and protein sequences are useful in wound healing, angiogenesis,  
 CC restenosis, atherosclerosis, inflammation, pulmonary diseases,  
 CC neurodegenerative diseases (such as Scrape, Alzheimer's disease, and  
 CC Creutzfeldt-Jakob disease) or viral infections. The heparanase coding  
 CC sequence is particularly useful in gene therapy  
 XX  
 XX Sequence 543 AA;  
 SQ

Query Match 100.0%; Score 82; DB 4; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
 |||||  
 Db 219 SWELGNEPNSFLKKA 233

RESULT 32  
 AAB86206  
 ID AAB86206 standard; protein; 543 AA.  
 XX  
 AC AAB86206;  
 XX  
 DT 24-AUG-2001 (first entry)  
 XX  
 DE Human heparanase inhibitor protein.  
 XX  
 KW Heparanase; inhibitor; cardiac insufficiency; cardiant; nephrotropic;  
 KW hepatotropic; veterinary medicine; congestive heart failure; dyspnoea;  
 KW primary cardiomyopathy; peripheral edema; pulmonary congestion;  
 KW hepatic congestion; hydrothorax; ascite; nocturia; human.  
 XX  
 OS Homo sapiens.  
 XX  
 FN DE19955803-A1.  
 XX  
 PD 23-MAY-2001.  
 XX  
 PF 19-NOV-1999; 99DE-01055803.  
 XX  
 PR 19-NOV-1999; 99DE-01055803.  
 XX  
 PA (KNOL ) KNOLL AG.  
 XX  
 PI Herr D, Hahn A, Laux V;  
 XX  
 DR WPI; 2001-368371/39.  
 DR N-PSDB; AAH20940.  
 XX  
 PT Treatment or prevention of cardiac insufficiency and related conditions,  
 PT e.g. pulmonary congestion and dyspnoea, comprises administration of  
 PT heparanase inhibitor.  
 XX  
 PS Disclosure; Page 11-13; 16pp; German.  
 XX

This invention describes a novel heparanase inhibitor which can be used  
 CC for the treatment or prevention of cardiac insufficiency and associated  
 CC indications, symptoms and/or malfunctions. The heparanase inhibitor of  
 CC the invention has cardiant, nephrotropic and hepatotropic activity. The  
 CC products of the invention can be used in human and veterinary medicine,  
 CC for the treatment or prevention of congestive heart failure e.g. primary  
 CC cardiomyopathy. Associated conditions treated or prevented with the  
 CC inhibitor are especially peripheral edemas, pulmonary and hepatic

CC congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g.  
 CC nocturia can also be treated. This sequence represents the human  
 CC heparanase protein described in the method of the invention  
 XX  
 SQ Sequence 543 AA;  
 Query Match 100.0%; Score 82; DB 4; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
 |||||  
 Db 219 SWELGNEPNSFLKKA 233

RESULT 33  
 AAB88361  
 ID AAB88361 standard; protein; 543 AA.  
 XX  
 AC AAB88361;  
 XX  
 DT 23-MAY-2001 (first entry)  
 XX  
 DE Human membrane or secretory protein clone PSEC0090.  
 XX  
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;  
 KW rheumatoid arthritis; diabetes.  
 XX  
 OS Homo sapiens.  
 XX  
 FN EP1067182-A2.  
 XX  
 PD 10-JAN-2001.  
 XX  
 PF 07-JUL-2000; 2000EP-00114090.  
 XX  
 PR 08-JUL-1999; 99JP-00194179.  
 PR 11-JAN-2000; 2000JP-00118775.  
 PR 02-MAY-2000; 2000JP-00183766.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
 XX  
 DR WPI; 2001-093989/11.  
 DR N-PSDB; AAF93788.  
 XX  
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in  
 PT gene therapy or as candidate target molecules in drug development.  
 XX  
 PS Claim 1; SEQ ID NO 90; 609pp + Sequence Listing; English.  
 XX  
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916  
 CC which encode human secretory or membrane proteins represented by AAB88317  
 CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and  
 CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the  
 CC invention. The invention also includes methods for the production of  
 CC antibodies directed against the proteins, and cDNA sequences, which can  
 CC be used in vaccines. The polynucleotide sequences can be used in gene  
 CC therapy. The polynucleotide sequences and the proteins they encode may be  
 CC used in the prevention, treatment and diagnosis of diseases associated  
 CC with inappropriate secretory protein/membrane protein expression. The  
 CC nucleic acids and complementary sequences may also be used as DNA probes  
 CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect  
 CC and quantitate the presence of similar nucleic acid sequences in samples.  
 CC They may also be used to study the expression and function of secretory  
 CC proteins/membrane polypeptides and their role in metabolism. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC against them and in assays to identify modulators (agonists and  
 CC antagonists) of expression and activity. The antibodies and antagonists  
 CC may also be used as therapeutic agents to down regulate expression and  
 CC activity. The antibodies may also be used as diagnostic agents for  
 CC detecting the presence of the polypeptides in samples (e.g. by enzyme



CC linked immunosorbant assay (ELISA). Examples of diseases which may be  
 XX treated include rheumatoid arthritis and diabetes

SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 4; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15

Db 219 SWELGNEPNSFLKKA 233

RESULT 34

ABB07813  
 ID ABB07813 standard; protein; 543 AA.

AC ABB07813;

DT 03-JUL-2002 (first entry)

DE Human heparanase sequence.

KW Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;  
 KW anti-protozoan; neuroprotective; heparin; human.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 1..35

FT /note= "signal peptide"

FT Protein 36..543

FT /note= "mature protein"

XX US2002034810-A1.

XX 21-MAR-2002.

XX 16-AUG-2001; 2001US-00930218.

XX 20-SEP-2000; 2000US-00666390.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;

XX WPI; 2002-338926/37.

XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful  
 PT to treat various heparin-related disorders and the signal peptide is  
 PT useful in production of membrane-targeted or secreted recombinant  
 PT proteins.

XX Disclosure; Fig 1a; 39pp; English.

XX The invention relates to an isolated avian and reptile nucleic acid,  
 CC encoding a polypeptide with heparanase catalytic activity. The signal  
 CC peptide of the nucleic acid can be used to express membrane-associated or  
 CC secreted proteins in heterologous expression systems. The encoded  
 CC polypeptides can be used to prevent tumour angiogenesis, metastasis and  
 CC invasion, and to intervene with pathologies associated with impaired  
 CC heparin-binding growth factors, cellular responses to heparin-binding  
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,  
 CC cellular susceptibility to viral, protozoa and bacterial infections or  
 CC disintegration of neurodegenerative plaques. The present sequence  
 CC represents a human heparanase protein sequence used in similarity studies

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 5; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15

Db 219 SWELGNEPNSFLKKA 233

RESULT 35

ADD18950

ID ADD18950 standard; protein; 543 AA.

XX AC ADD18950;

XX DT 15-JAN-2004 (first entry)

XX Human disease related protein SeqID439.

XX human; disease state; cytostatic; antiinflammatory; ophthalmological;  
 KW antiarteriosclerotic; vulnery; gene therapy;  
 KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;  
 KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;  
 KW glucose transportation; catecholamine synthesis; iron transport;  
 KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;  
 KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;  
 KW inflammatory condition; wound healing.

OS Homo sapiens.

XX WO2003018621-A2.

XX PD 06-MAR-2003.

XX 23-AUG-2002; 2002WO-GB003892.

XX 23-AUG-2001; 2001GB-00020558.

XX 05-OCT-2001; 2001GB-00024037.

XX (OXFO-) OXFORD BIOMEDICA UK LTD.

XX Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;

XX WPI; 2003-290046/28.

XX DR N-PSDB; ADD18951.

XX New substantially purified polypeptide, useful for diagnosing or treating  
 PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion  
 PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or  
 PT wound healing.

XX Claim 25; SEQ ID NO 439; 424pp; English.

XX This invention relates to novel human genes and gene product which are  
 CC implicated in certain disease states. Compounds which modulate the  
 CC proteins of the invention may have cytostatic, antiinflammatory,  
 CC ophthalmological, antiarteriosclerotic or vulnery activities. The  
 CC sequences of the invention may be useful for gene therapy. The invention  
 CC may be useful for diagnosing or treating a hypoxia-regulated condition,  
 CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,  
 CC erythropoiesis, or the biological response to hypoxia. Conditions  
 CC including processes such as glycolysis, gluconeogenesis, glucose  
 CC transportation, catecholamine synthesis, iron transport or nitric oxide  
 CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion  
 CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,  
 CC inflammatory conditions or wound healing. The present sequence is that of  
 CC a disease related protein of the invention.

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 7; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15

|||||

Db 219 SWELGNEPNSFLKKA 233

RESULT 36  
ADG88800  
ID ADG88800 standard; protein; 543 AA.  
XX  
AC ADG88800;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Human hpa protein.  
XX  
KW Wound healing; heparanase; ulcer; burn; laceration; surgical incision;  
KW necrosis; pressure wound; diabetic ulcer; angiogenesis; human; therapy.  
XX  
OS Homo sapiens.  
XX  
PN US2003161823-A1.  
XX  
PD 28-AUG-2003.  
XX  
PF 14-JAN-2003; 2003US-00341582.  
XX  
PR 31-AUG-1998; 98WO-US017954.  
XX  
PR 01-MAR-1999; 99US-00258892.  
XX  
PR 06-FEB-2001; 2001US-00776874.  
XX  
PR 05-SEP-2001; 2001WO-IL000830.  
XX  
PR 19-NOV-2001; 2001US-00988113.  
XX  
PA (ILAN/) ILAN N.  
PA (VL0D/) VL0DAVSKY I.  
PA (YACO/) YACOBY-ZEEVI O.  
PA (PECK/) PECKER I.  
PA (FEIN/) FEINSTEIN E.  
XX  
PI Ilan N, Vlodavsky I, Yacoby-Zeevi O, Pecker I, Feinstein E;  
XX  
DR WPI; 2003-897910/82.  
DR N-PSDB; ADG88799, ADG88801, ADG88832.  
XX  
PT Composition for treating a wound comprising recombinant heparanase is  
PT useful to induce or accelerate wound healing and induce or accelerate  
PT angiogenesis.  
XX  
PS Claim 2; SEQ ID NO 10; 143pp; English.  
XX  
CC The present invention relates to methods and compositions for inducing  
CC and/or accelerating wound healing via the catalytic activity of  
CC heparanase. The invention is used to induce or accelerate a healing  
CC process, particularly of an ulcer, burn, laceration, surgical incision,  
CC necrosis, pressure wound, diabetic ulcer and to induce or accelerate  
CC angiogenesis. The present sequence is human hpa protein.  
XX  
SQ Sequence 543 AA;  
Query Match 100.0%; Score 82; DB 7; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 SWELGNEPNSFLKKA 15  
Db 219 SWELGNEPNSFLKKA 233  
RESULT 37  
ADL16379  
ID ADL16379 standard; protein; 543 AA.  
XX  
AC ADL16379;  
XX  
DT 06-MAY-2004 (first entry)  
XX

DE Human heparanase partial protein.  
XX  
KW Human; heparanase; heparanase-dependent cancer; cancer;  
KW autoimmune reaction; inflammation; chromosome 4; enzyme.  
XX  
OS Homo sapiens.  
XX  
PN US2003236215-A1.  
XX  
PD 25-DEC-2003.  
XX  
PF 09-JUN-2003; 2003US-00456573.  
XX  
PR 31-AUG-1998; 98WO-US017954.  
PR 01-MAR-1999; 99US-00258892.  
PR 08-NOV-1999; 99US-00435739.  
XX  
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
XX  
PI Pecker I, Vlodavsky I, Feinstein E;  
XX  
DR WPI; 2004-070610/07.  
XX  
CC New antisense oligonucleotide hybridizable with a polynucleotide encoding  
CC a polypeptide with heparanase activity, useful for treating diseases such  
CC as cancer and autoimmune disorders.  
XX  
PS Claim 3; SEQ ID NO 10; 108pp; English.  
XX  
CC The invention relates to an antisense oligonucleotide (ASO) comprising a  
CC polynucleotide or a polynucleotide analogue of at least 10 bases being  
CC hybridizable in vivo, under physiological conditions, with a portion of  
CC a polynucleotide strand encoding a polypeptide having heparanase  
CC catalytic activity. Also included are a method of in vivo downregulating  
CC heparanase activity (comprising administering the ASO in vivo), a method  
CC of treating a subject suffering from a pathological condition  
CC (characterised by heparanase activity, comprising administering ASO to  
CC the subject), a pharmaceutical composition comprising the ASO and a  
CC carrier, an antisense nucleic acid construct (comprising a promoter  
CC sequence and a polynucleotide sequence directing the synthesis of an  
CC antisense RNA sequence of at least 10 bases being hybridisable in vivo,  
CC under physiological conditions, with a polynucleotide strand encoding a  
CC polypeptide having heparanase catalytic activity), a method of in vivo  
CC downregulating heparanase activity (comprising administering in vivo the  
CC antisense nucleic acid construct), a pharmaceutical composition  
CC comprising the antisense nucleic acid construct and a carrier, and an  
CC antisense oligonucleotide comprising a polynucleotide or a polynucleotide  
CC analogue of at least 10 bases being hybridisable in vivo, under  
CC physiological conditions, with a portion of a polynucleotide strand being  
CC characterised by forming at least a portion of an untranslated region  
CC (UTR) for a polynucleotide strand encoding a polypeptide having  
CC heparanase catalytic activity. The methods and compositions of the  
CC present invention are useful for the prevention and/or treatment of  
CC diseases or conditions associated with aberrant heparanase activity, such  
CC as heparanase-dependent cancer, cancer, autoimmune reaction and  
CC inflammation. The gene for human heparanase is located on chromosome 4.  
XX  
CC The present sequence is a human heparanase protein.  
XX  
SQ Sequence 543 AA;  
Query Match 100.0%; Score 82; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 SWELGNEPNSFLKKA 15  
Db 219 SWELGNEPNSFLKKA 233  
RESULT 38  
ADK52086  
ID ADK52086 standard; protein; 543 AA.

XX AC ADK52086;  
XX DT 20-MAY-2004 (first entry)  
XX DE Human atopic dermatitis/psoriasis-associated protein #1.  
XX KW Human; atopic dermatitis; psoriasis; dermatological; anti-inflammatory;  
XX OS antipsoriatic; rash.  
XX PN Homo sapiens.  
XX WO2004016785-A1.  
XX PD 26-FEB-2004.  
XX PF 06-AUG-2003; 2003WO-JP009999.  
XX PR 06-AUG-2002; 2002JP-00229319.  
XX PR 14-MAY-2003; 2003JP-00136544.  
XX PA (GENO-) GENOX RES INC.  
XX PA (UYJU-) UNIV JUNTENDO.  
XX PI Itoh M, Ogawa K, Shinagawa A, Sudo H, Ogawa H, Ra C;  
XX PI Mitsuishi K;  
XX DR WPI; 2004-214514/20.  
XX DR N-PSDB; ADK51968.  
XX PT Detecting atopic dermatitis or psoriasis comprises assaying levels of  
XX PT expression of an indicator gene at a rash site and non-rash site of a  
XX PT person with atopic dermatitis or psoriasis.  
XX PS Example 2; SEQ ID NO 119; 484pp; Japanese.  
XX CC The invention relates to detecting atopic dermatitis or psoriasis  
XX CC comprising assaying the levels of expression of an indicator gene at a  
XX CC rash site and non-rash site of a person with atopic dermatitis or  
XX CC psoriasis, comparing these levels with those of a healthy person, and  
XX CC determining that if the levels of indicators are higher or lower, then  
XX CC this indicates the disease. Also included are a reagent for detecting  
XX CC atopic dermatitis or psoriasis, a kit for screening for treatments, a  
XX CC transgenic non human vertebrate animal models for the diseases, an agent  
XX CC for inducing the diseases in mice and a DNA chip for assaying for the  
XX CC indicator genes. The method is used for treatment, detection and animal  
XX CC models for research of atopic dermatitis and psoriasis. The present  
XX CC sequence is a protein encoded by an indicator gene of the invention.  
XX SQ Sequence 543 AA;  
Query Match 100.0%; Score 82; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SWELGNEPNSFLKKA 15  
Db 219 SWELGNEPNSFLKKA 233  
RESULT 39  
ADM48716  
ID ADM48716 standard; protein; 543 AA.  
XX AC ADM48716;  
XX DT 03-JUN-2004 (first entry)  
XX DE Human hpa protein #1.  
XX KW Transgenic animal; heparanase; cancer; viral infection; restenosis;  
XX KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;  
XX KW human.

XX OS Homo sapiens.  
XX PN US2003217375-A1.  
XX PD 20-NOV-2003.  
XX PF 24-FEB-2003; 2003US-00371218.  
XX PR 31-AUG-1998; 98WO-US017954.  
XX PR 01-MAR-1999; 99US-00258892.  
XX PR 06-FEB-2001; 2001US-00776874.  
XX PR 19-NOV-2001; 2001US-00988113.  
XX PA (ZCHA/) ZCHARIA E.  
XX PA (VLOD/) VLODAVSKY I.  
XX PA (METZ/) METZGER S.  
XX PA (PECK/) PECKER I.  
XX PA (ILAN/) ILAN N.  
XX PA (CHAJ/) CHAJEK-SHAUL T.  
XX PA (GOLD/) GOLDSCHMIDT O.  
XX PI Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;  
XX PI Chajek-Shaul T, Goldshmidt O;  
XX DR WPI; 2004-021918/02.  
XX DR N-PSDB; ADM48715, ADM48717.  
XX PT New transgenic non-human animal expressing heparinase, useful as models  
XX PT for human disease, such as cancers, viral infection, neurodegenerative  
XX PT diseases, restenosis, atherosclerosis and pulmonary disorders.  
XX PS Example 1; SEQ ID NO 10; 106pp; English.  
XX CC The present invention relates to a transgenic non-human animal whose  
XX CC genome comprises an exogenous polynucleotide sequence, including a  
XX CC promoter active in tissues of the non-human, a region encoding a human  
XX CC heparanase, where the promoter and the region encoding human heparanase  
XX CC are operably linked in the exogenous polynucleotide such that human  
XX CC heparanase is expressed in at least a portion of the cells of the non-  
XX CC human animal. The methods and compositions of the present invention are  
XX CC useful for the production of transgenic animals expressing heparanase, to  
XX CC be used as models for human diseases such as cancers, viral infection,  
XX CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary  
XX CC disorders. The present sequence is human hpa protein used in the  
XX CC exemplification of the invention.  
XX SQ Sequence 543 AA;  
Query Match 100.0%; Score 82; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SWELGNEPNSFLKKA 15  
Db 219 SWELGNEPNSFLKKA 233  
RESULT 40  
ADM48759  
ID ADM48759 standard; protein; 543 AA.  
XX AC ADM48759;  
XX DT 03-JUN-2004 (first entry)  
XX DE Human hpa protein #2.  
XX KW Transgenic animal; heparanase; cancer; viral infection; restenosis;  
XX KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;  
XX KW human.  
XX OS Homo sapiens.

XX PN US2003217375-A1.  
XX PD 20-NOV-2003.  
XX PF 24-FEB-2003; 2003US-00371218.  
XX PR 31-AUG-1998; 98WO-US017954.  
XX PR 01-MAR-1999; 99US-00258892.  
XX PR 06-FEB-2001; 2001US-00776874.  
XX PR 19-NOV-2001; 2001US-00988113.  
XX PA (ZCHA/) ZCHARIA E.  
XX PA (VL0D/) VLODAVSKY I.  
XX PA (METZ/) METZGER S.  
XX PA (PECK/) PECKER I.  
XX PA (ILAN/) ILAN N.  
XX PA (CHAJ/) CHAJEK-SHAUL T.  
XX PA (GOLD/) GOLDSHMIDT O.  
XX PI Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;  
XX PI Chajek-Shaul T, Goldshmidt O;  
XX DR WPI; 2004-021918/02.  
XX DR N-PSDB; ADM48748.  
XX PS New transgenic non-human animal expressing heparinase, useful as models  
XX PT for human disease, such as cancer, viral infection, neurodegenerative  
XX PT diseases, restenosis, atherosclerosis and pulmonary disorders.  
XX PS Example 10; Fig 16; 106pp; English.  
XX CC The present invention relates to a transgenic non-human animal whose  
XX CC genome comprises an exogenous polynucleotide sequence, including a  
XX CC promoter active in tissues of the non-human, a region encoding a human  
XX CC heparanase, where the promoter and the region encoding human heparanase  
XX CC are operably linked in the exogenous polynucleotide such that human  
XX CC heparanase is expressed in at least a portion of the cells of the non-  
XX CC human animal. The methods and compositions of the present invention are  
XX CC useful for the production of transgenic animals expressing heparanase, to  
XX CC be used as models for human diseases such as cancer, viral infection,  
XX CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary  
XX CC disorders. The present sequence is human hpa protein used in the  
XX CC exemplification of the invention.  
XX SQ Sequence 543 AA;  
Query Match 100.0%; Score 82; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 SWELGNEPNSFLKKA 15  
Db 219 SWELGNEPNSFLKKA 233  
RESULT 41  
ID ADN05074 standard; protein; 543 AA.  
XX AC ADN05074;  
XX AC ADN05074;  
XX DT 01-JUL-2004 (first entry)  
XX DE Antipsoriatic protein sequence #716.  
XX KW antipsoriatic; gene therapy; psoriasis; diagnosis.  
XX OS Homo sapiens.  
XX PN WO2004028479-A2.  
XX PD 08-APR-2004.  
Qy 1 SWELGNEPNSFLKKA 15  
Db 219 SWELGNEPNSFLKKA 233  
RESULT 41  
ID ADN05074 standard; protein; 543 AA.  
XX AC ADN05074;  
XX AC ADN05074;  
XX DT 01-JUL-2004 (first entry)  
XX DE Antipsoriatic protein sequence #716.  
XX KW antipsoriatic; gene therapy; psoriasis; diagnosis.  
XX OS Homo sapiens.  
XX PN WO2004028479-A2.  
XX PD 08-APR-2004.

XX PF 25-SEP-2003; 2003WO-US030907.  
XX PR 25-SEP-2002; 2002US-0414006P.  
XX PA (GETH ) GENENTECH INC.  
XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;  
XX PI Wu TD;  
XX XX WPI; 2004-305105/28.  
XX DR N-PSDB; ADN05073.  
XX PT New PRO nucleic acid or polypeptide, useful for preparing a  
XX PT pharmaceutical composition for diagnosing or treating psoriasis in a  
XX PT mammal.  
XX PS Claim 9; SEQ ID NO 1468; 3069pp; English.  
XX CC The invention relates to novel polynucleotide and polypeptides for  
XX CC treating psoriasis or a sequence having at least 80% identity to the  
XX CC above sequences. The nucleic acid is useful for preparing a composition  
XX CC for diagnosing or treating psoriasis in a mammal. This sequence  
XX CC corresponds to one of the polypeptides of the invention.  
XX SQ Sequence 543 AA;  
Query Match 100.0%; Score 82; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 SWELGNEPNSFLKKA 15  
Db 219 SWELGNEPNSFLKKA 233  
RESULT 42  
ID ADN04902 standard; protein; 543 AA.  
XX AC ADN04902;  
XX AC ADN04902;  
XX DT 01-JUL-2004 (first entry)  
XX DE Antipsoriatic protein sequence #631.  
XX KW antipsoriatic; gene therapy; psoriasis; diagnosis.  
XX OS Homo sapiens.  
XX PN WO2004028479-A2.  
XX PD 08-APR-2004.  
XX PF 25-SEP-2003; 2003WO-US030907.  
XX PR 25-SEP-2002; 2002US-0414006P.  
XX PA (GETH ) GENENTECH INC.  
XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;  
XX PI Wu TD;  
XX XX WPI; 2004-305105/28.  
XX DR N-PSDB; ADN04901.  
XX PT New PRO nucleic acid or polypeptide, useful for preparing a  
XX PT pharmaceutical composition for diagnosing or treating psoriasis in a  
XX PT mammal.  
XX PS Claim 9; SEQ ID NO 1296; 3069pp; English.  
XX CC The invention relates to novel polynucleotide and polypeptides for

CC treating psoriasis or a sequence having at least 80% identity to the  
 CC above sequences. The nucleic acid is useful for preparing a composition  
 CC for diagnosing or treating psoriasis in a mammal. This sequence  
 CC corresponds to one of the polypeptides of the invention.

XX  
 SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
 |||||  
 DB 219 SWELGNEPNSFLKKA 233

RESULT 43  
 ADO63831  
 ID ADO63831 standard; protein; 543 AA.  
 XX  
 AC ADO63831;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Human heparanase mutant E378A.

XX Human; heparanase; heparanase-derived protein; heparanase mutant;  
 KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;  
 KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;  
 KW vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnery; mutant; mutein; enzyme.

XX Homo sapiens.  
 OS Synthetic.

XX  
 FH Key Location/Qualifiers  
 FT Active-site 225 /note= "Active site proton donor"  
 FT Active-site 343 /note= "Active site nucleophile"  
 FT Misc-difference 378 /note= "Ala replaces wild-type Glu"  
 FT  
 FT  
 XX WO2004048558-A2.  
 XX 10-JUN-2004.  
 XX 24-NOV-2003; 2003WO-IL000989.  
 XX 24-NOV-2002; 2002IL-00153059.  
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
 XX WPI; 2004-450373/42.  
 XX New nucleic acid construct comprising heparanase-derived polypeptide,  
 XX useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 XX Thrombasthenia, or Bernard-Soulier syndrome.

XX  
 XX Example 4; Page; 128pp; English.

XX The invention relates to nucleic acid constructs comprising a nucleic  
 CC acid encoding a heparanase-derived protein which lacks heparanase  
 CC endoglycosidase catalytic activity but which retains its cell-cell and  
 CC cell-matrix adhesion properties. The constructs of the invention  
 CC optionally further comprise operably linked regulatory elements. The  
 CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the

CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
 CC host cells are useful in preparing a tissue sealant composition for  
 CC sealing injuries, reducing the loss of blood, accelerating the healing  
 CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialisation of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be used in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte  
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be used to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents a human  
 CC heparanase mutant E378A created in an example of the invention which  
 CC retains its heparanase catalytic activity. The present sequence is not  
 CC shown in the invention, but is derived from the protein sequence of  
 CC GenBank accession number AF144325 and the information provided on page  
 CC 70.

XX  
 SQ Sequence 543 AA;

Query Match 100.0%; Score 82; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
 |||||  
 DB 219 SWELGNEPNSFLKKA 233

RESULT 44  
 ADO63823  
 ID ADO63823 standard; protein; 543 AA.  
 XX  
 AC ADO63823;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Human heparanase mutant E343A, SEQ ID:8.

XX Human; heparanase; heparanase-derived protein; heparanase mutant;  
 KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;  
 KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;  
 KW vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnery; mutant; mutein.

XX Homo sapiens.  
 OS Synthetic.

XX  
 FH Key Location/Qualifiers  
 FT Active-site 225 /note= "Active site proton donor"  
 FT Misc-difference 343 /note= "Ala replaces wild-type Glu (active site  
 XX nucleophile)"  
 XX WO2004048558-A2.  
 XX 10-JUN-2004.  
 XX 24-NOV-2003; 2003WO-IL000989.  
 XX 24-NOV-2002; 2002IL-00153059.  
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;

XX WPI; 2004-450373/42.  
 DR N-PSDB; ADO63817.  
 XX  
 FT New nucleic acid construct comprising heparanase-derived polypeptide,  
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 PT thrombasthenia, or Bernard-Soulier syndrome.  
 XX  
 XX  
 PS Claim 9; SEQ ID NO 8; 128pp; English.  
 XX  
 CC The invention relates to nucleic acid constructs comprising a nucleic  
 CC acid encoding a heparanase-derived protein which lacks heparanase  
 CC endoglycosidase catalytic activity but which retains its cell-cell and  
 CC cell-matrix adhesion properties. The constructs of the invention  
 CC optionally further comprise operably linked regulatory elements. The  
 CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the  
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
 CC host cells are useful in preparing a tissue sealant composition for  
 CC sealing injuries, reducing the loss of blood, accelerating the healing  
 CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialisation of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be used in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte  
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be used to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents the human  
 CC heparanase mutant E343A.  
 XX  
 SQ Sequence 543 AA;  
 Query Match 100.0%; Score 82; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SWELGNPNFLKKA 15  
 |||||  
 Db 219 SWELGNPNFLKKA 233  
 RESULT 45  
 ADO63832  
 ID ADO63832 standard; protein; 543 AA.  
 XX  
 AC ADO63832;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Human heparanase mutant E396A.  
 XX  
 KW Human; heparanase; heparanase-derived protein; heparanase mutant;  
 KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;  
 KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;  
 KW vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnerable; mutant; mutagen; enzyme.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PH Key Location/Qualifiers  
 FT Active-site 225  
 FT /note= "Active site proton donor"  
 FT Active-site 343

FT  
 FT Misc-difference 396  
 FT /note= "Active site nucleophile"  
 XX  
 PN W02004048558-A2.  
 PD 10-JUN-2004.  
 XX  
 XX 24-NOV-2003; 2003WO-11000989.  
 XX  
 XX 24-NOV-2002; 2002IL-00153059.  
 PR  
 PR (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA  
 XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
 PI WPI; 2004-450373/42.  
 XX  
 DR New nucleic acid construct comprising heparanase-derived polypeptide,  
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 PT thrombasthenia, or Bernard-Soulier syndrome.  
 XX  
 PS Example 4; Page; 128pp; English.  
 CC The invention relates to nucleic acid constructs comprising a nucleic  
 CC acid encoding a heparanase-derived protein which lacks heparanase  
 CC endoglycosidase catalytic activity but which retains its cell-cell and  
 CC cell-matrix adhesion properties. The constructs of the invention  
 CC optionally further comprise operably linked regulatory elements. The  
 CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the  
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
 CC host cells are useful in preparing a tissue sealant composition for  
 CC sealing injuries, reducing the loss of blood, accelerating the healing  
 CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialisation of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be used in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte  
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be used to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents a human  
 CC heparanase mutant E378A created in an example of the invention which  
 CC retains its heparanase catalytic activity. The present sequence is not  
 CC shown in the invention, but is derived from the protein sequence of  
 CC GenBank accession number AF144325 and the information provided on page  
 CC 70.  
 XX  
 SQ Sequence 543 AA;  
 Query Match 100.0%; Score 82; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SWELGNPNFLKKA 15  
 |||||  
 Db 219 SWELGNPNFLKKA 233  
 RESULT 46  
 ADO80372  
 ID ADO80372 standard; protein; 543 AA.  
 XX  
 AC ADO80372;  
 XX  
 DT 21-OCT-2004 (first entry)

```

XX DE Heparanase protein.
XX DE
XX KW cytostatic; epidermal growth factor receptor modulator; identification;
XX KW therapeutic response; cancer; EGFR; biomarker.
XX OS Homo sapiens.
XX PN WO2004063709-A2.
XX PD 29-JUL-2004.
XX XX
XX PF 08-JAN-2004; 2004WO-US000368.
XX XX
XX PR 08-JAN-2003; 2003US-0438735P.
XX XX
XX PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX XX
XX PI Amler LC, Januario T;
XX XX
XX DR WPI; 2004-544114/52.
XX DR N-PSDB; ADQ80253.
XX XX
XX PT Identifying a mammal that will respond therapeutically to a method of
XX PT treating cancer comprises comparing the level of a biomarker in a mammal
XX PT before and after exposure to an epidermal growth factor receptor (EGFR)
XX PT modulator.
XX XX
XX PS Disclosure; SEQ ID NO 144; 520pp; English.
XX XX
XX CC The invention relates to a method of identifying a mammal that will
XX CC respond therapeutically to a method of treating cancer by administering
XX CC an epidermal growth factor receptor (EGFR) modulator by comparing the
XX CC level of a biomarker in a mammal before and after exposure to an EGFR
XX CC modulator. The method comprises: (a) measuring, in the mammal, the level
XX CC of at least one biomarker identified in the specification; (b) exposing
XX CC the mammal to the EGFR modulator; and (c) measuring in the mammal the
XX CC level of the biomarker, where a difference in the level in step (c)
XX CC compared to step (a) indicates that the mammal will respond
XX CC therapeutically to the method of treating cancer. The method and
XX CC biomarkers are useful for identifying a mammal that will respond
XX CC therapeutically to a method of treating cancer by administering an
XX CC epidermal growth factor receptor (EGFR) modulator. This sequence
XX CC corresponds to one of the biomarkers whose levels of expression is
XX CC measured in the method of the invention.
XX SQ Sequence 543 AA;
XX
XX Query Match 100.0%; Score 82; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-05;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 SWELGNEPNSFLKKA 15
XX |||||
XX Db 219 SWELGNEPNSFLKKA 233
XX
XX RESULT 47
XX ADR88210
XX ID ADR88210 standard; protein; 543 AA.
XX AC
XX AC ADR88210;
XX XX
XX DT 18-NOV-2004 (first entry)
XX XX
XX DE Human preproheparanase.
XX XX
XX CC Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;
XX KW autoimmune disorder; cancer; angiogenesis; metastatic disease;
XX KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;
XX KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;
XX KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;
XX KW human; heparanase; enzyme.

```

---

```

XX OS Homo sapiens.
XX FH Key
XX FT Peptide
XX FT Protein
XX FT Region
XX FT Domain
XX FT Region
XX FT Domain
XX FT Active-site
XX FT Binding-site
XX FT Domain
XX FT Domain
XX FT Active-site
XX FT Binding-site
XX FT Domain
XX US2004170631-A1.
XX 02-SEP-2004.
XX 28-NOV-2003; 2003US-00722502.
XX 02-SEP-1997; 97US-00922170.
XX 01-MAY-1998; 98US-00071739.
XX 04-NOV-1998; 98US-00186200.
XX 19-FEB-2003; 2003US-00368044.
XX 22-AUG-2003; 2003US-00645659.
XX (YACO/) YACOBY-ZEEVI O.
XX (PERE/) PERETZ T.
XX (MIRO/) MIRON D.
XX (SHLO/) SHLOMI Y.
XX (PECK/) PECKER I.
XX (AYAL/) AYAL-HERSHKOVITZ M.
XX (FEIN/) FEINSTEIN E.
XX (VGL/) VAN GELDER J M.
XX (VLOD/) VLODAVSKY I.
XX (FRIE/) FRIEDMANN Y.
XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;
XX Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;
XX Friedmann Y;
XX WPI; 2004-625084/60.
XX Targeted drug delivery to a heparanase-expressing tissue of a patient,
XX useful for treating heparanase-associated conditions such as inflammation
XX or cancer, comprises administering a drug and an anti-heparanase antibody
XX complex.
XX Claim 2; SEQ ID NO 4; 58pp; English.
XX The invention relates to a method of targeted drug delivery to a tissue
XX of a patient, the tissue expressing heparanase. The method comprises
XX providing a complex of a drug directly or indirectly linked to an anti-
XX heparanase antibody, and administering the complex to the patient. In the
XX targeted drug delivery, the antibody comprises an antibody or its portion
XX capable of specifically binding to at least one epitope of a heparanase

```



CC protein. The composition and methods of the invention are useful for  
 CC diagnosing, preventing or treating conditions associated with heparanase  
 CC catalytic activity (e.g. an inflammatory disorder, wound, scar,  
 CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell  
 CC proliferation, invasion of circulating tumour cells and metastatic  
 CC disease), for purifying heparanase, or for developing drugs for those  
 CC heparanase-associated conditions. The vasculopathy is atherosclerosis,  
 CC restenosis or aneurysm. The cancerous condition is a solid cancer or a  
 CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy  
 CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous  
 CC leukaemia (AML), chronic lymphocytic leukaemia (CLL), chronic myelogenous  
 CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,  
 CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and  
 CC multiple myeloma. The solid cancer is selected from tumours in lip and  
 CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,  
 CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,  
 CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of  
 CC Vater, exocrine pancreas, lung, pleural mesothelium, soft tissue  
 CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,  
 CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic  
 CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary  
 CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,  
 CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,  
 CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,  
 CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's  
 CC sarcoma. The present sequence is human preproheparanase.

XX Sequence 543 AA;

Query Match 100.0%; Score 82; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
 |||||  
 Db 219 SWELGNEPNSFLKKA 233

RESULT 48  
 ADP25079  
 ID ADP25079 standard; protein; 543 AA.

XX ADP25079;

XX 18-NOV-2004 (first entry)

XX PRO polypeptide SEQ ID NO:2257.

XX PRO; antiinflammatory; antiarthritic; antirheumatic; immunosuppressive;  
 KW osteopathic; antidiabetic; dermatological; antipsoriatic; antiallergic;  
 KW antiasthmatic; hepatotropic; respiratory; gene therapy; immune system.

XX Unidentified.

XX WO2004041170-A2.

XX 21-MAY-2004.

XX 30-OCT-2003; 2003WO-US034312.

XX 01-NOV-2002; 2002US-0423394P.

XX (GETH ) GENENTECH INC.

XX Clark H, Schoenfeld J, Van Lookeren M, Williams PM, Wood WI;  
 PI Wu TD;

XX WPI: 2004-419628/39.

XX N-FSDB; ADP25078.

XX New PRO polypeptides and polynucleotides, useful for treating e.g.  
 PT erythematous, rheumatoid arthritis, diabetes mellitus, immune-mediated  
 PT renal disease, or demyelinating diseases of the central or peripheral

PT nervous system.

XX Claim 7; SEQ ID NO 2257; 2940pp; English.

XX The invention relates to a novel isolated nucleic acid and the PRO  
 CC polypeptide encoded by it. A protein of the invention has  
 CC antiinflammatory, antiarthritic, antirheumatic, immunosuppressive,  
 CC osteopathic, antidiabetic, dermatological, antipsoriatic, antiallergic,  
 CC antiasthmatic, hepatotropic, and respiratory activity. A polynucleotide  
 CC of the invention may have a use in gene therapy. The PRO polypeptide, its  
 CC agonist, antagonist, or antibody that specifically binds to the  
 CC polypeptide is useful for treating an immune related disorder such as  
 CC systemic lupus erythematosus, rheumatoid arthritis, osteoarthritis,  
 CC juvenile chronic arthritis, a spondyloarthropathy, systemic sclerosis, an  
 CC idiopathic inflammatory myopathy, Sjogren's syndrome, systemic  
 CC vasculitis, sarcoidosis, autoimmune haemolytic anaemia, autoimmune  
 CC thrombocytopenia, thyroiditis, diabetes mellitus, immune-mediated renal  
 CC disease, a demyelinating disease of the central or peripheral nervous  
 CC system, idiopathic demyelinating polyneuropathy, Guillain-Barre syndrome,  
 CC a chronic inflammatory demyelinating polyneuropathy, a hepatobiliary  
 CC disease, infectious or autoimmune chronic active hepatitis, primary  
 CC biliary cirrhosis, granulomatous hepatitis, sclerosing cholangitis,  
 CC inflammatory bowel disease, gluten-sensitive enteropathy, Whipple's  
 CC disease, an autoimmune or immune-mediated skin disease, a bullous skin  
 CC disease, erythema multiforme, contact dermatitis, psoriasis, an allergic  
 CC disease, asthma, allergic rhinitis, atopic dermatitis, food  
 CC hypersensitivity, urticaria, an immunologic disease of the lung,  
 CC eosinophilic pneumonia, idiopathic pulmonary fibrosis, hypersensitivity  
 CC pneumonitis, a transplantation associated disease, graft rejection or  
 CC graft-versus-host disease. The present sequence represents a PRO protein  
 CC of the invention.

XX Sequence 543 AA;

Query Match 100.0%; Score 82; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
 |||||  
 Db 219 SWELGNEPNSFLKKA 233

RESULT 49

ADT78177

ID ADT78177 standard; protein; 543 AA.

XX ADT78177;

XX 13-JAN-2005 (first entry)

XX Human heparanase protein.

XX Antibody; epitope; heparanase; pathological condition; angiogenesis;  
 KW cell proliferation; cancerous condition; tumour cell invasion;  
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;  
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;  
 KW cytostatic; antiinflammatory; vulnery; antiarteriosclerotic;  
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Binding-site 157..162

FT Binding-site /note= "Putative heparin binding site"

FT Binding-site 271..277

FT Binding-site /note= "Putative heparin binding site"

FT Binding-site 426..433

FT Binding-site /note= "Putative heparin binding site"

XX US20040213789-A1.

XX 28-OCT-2004.

PD

XX 22-AUG-2003; 2003US-00645659.  
XX  
XX 02-SEP-1997; 97US-00922170.  
PR 01-MAY-1998; 98US-00071739.  
PR 04-NOV-1998; 98US-00186200.  
PR 19-FEB-2003; 2003US-00368044.  
XX  
XX (YACO/) YACOBY-ZEEVI O.  
PA (PERE/) PERETZ T.  
PA (MIRO/) MIRON D.  
PA (SHLO/) SHLOMI Y.  
PA (PECK/) PECKER I.  
PA (AYAL/) AYAL-HERSHKOVITZ M.  
PA (FEIN/) FEINSTEIN E.  
PA (GELD/) GELDER J M V.  
PA (VLOB/) VLODAVSKY I.  
PA (FRIE/) FRIEDMANN Y.  
XX  
XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
PI Ayal-Hershkovitz M, Feinstein E, Gelder JMV, Vlodavsky I;  
PI Friedmann Y;  
XX  
XX WPI; 2004-774790/76.  
XX  
XX New neutralizing monoclonal anti-heparanase antibodies, useful for  
PT detecting, treating or preventing cancer, inflammatory or autoimmune  
PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.  
XX  
XX Claim 5; SEQ ID NO 4; 68pp; English.  
XX  
XX The invention relates to an isolated antibody or antibody portion capable  
CC of specifically binding to or elicited by at least one epitope of a  
CC heparanase protein, where the heparanase protein is at least 60%  
CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and  
CC where at least one epitope comprises a sequence at least 70% homologous  
CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)  
CC a hybridoma cell line comprising a cell line for producing the monoclonal  
CC antibody, (b) a method for detecting, treating or preventing a  
CC pathological condition or a heparanase-related disorder or condition in a  
CC subject, (c) a method for monitoring the state of a heparanase-related  
CC disorder or condition in a subject, and (d) a pharmaceutical composition  
CC comprising the isolated anti-heparanase antibody or antibody portion and  
CC a pharmaceutical carrier. The antibody, methods, and composition are  
CC useful for detecting, treating, preventing or monitoring a pathological  
CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition  
CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,  
CC or prostate cancer), minor cell proliferation, invasion of circulating  
CC tumour cells, or a metastatic disease, or a heparanase-related disorder  
CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy  
CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or  
CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,  
CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell  
CC carcinoma) in a mammal. This sequence represents human heparanase.  
XX  
XX Sequence 543 AA;  
SQ  
Query Match 100.0%; Score 82; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SWELGNPNFLKKA 15  
DB 219 SWELGNPNFLKKA 233  
RESULT 50  
ADY27036  
ID ADY27036 standard; protein; 543 AA.  
XX  
XX AC ADY27036;  
XX  
XX DT 05-MAY-2005 (first entry)

XX Human heparanase protein.  
DE  
XX  
XX Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
KW neurological disease; viral infection; infection; cytostatic;  
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
KW protease; enzyme; enzyme purification.  
XX  
XX Homo sapiens.  
OS  
XX WO2005016227-A2.  
PN  
XX 24-FEB-2005.  
PD  
XX 12-AUG-2004; 2004WO-IL000744.  
PF  
XX 14-AUG-2003; 2003US-0494800P.  
PR  
XX 12-JAN-2004; 2004US-0535492P.  
PR  
XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.  
PA  
XX Van-Gelder JM, Miron D;  
PI  
XX WPI; 2005-182203/19.  
DR  
XX  
XX Regulating heparanase activity, useful for treating heparanase-associated  
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
PT neurological diseases or viral diseases) comprises modulating heparanase  
PT activation.  
XX  
XX Disclosure; SEQ ID NO 8; 21pp; English.  
PS  
XX The invention relates to a method of regulating heparanase activity in a  
CC tissue or regulating a biological process depending at least in part on  
CC heparanase activity comprising modulating heparanase activation. The  
CC invention also relates to methods of treating a heparanase- or heparin  
CC binding protein-associated disease or disorder in a subject, a  
CC pharmaceutical composition for use in the treatment of a heparanase-  
CC associated disease or disorder comprising a therapeutic amount of an  
CC agent capable of modulating heparanase activation and a pharmaceutical  
CC carrier or diluent, a method of identifying a protease activator of  
CC heparanase, a protease substrate mimetic comprising a peptide  
CC representing a subset or all substrate residues or cleavage sites of  
CC human heparanase or an equivalent non-human heparanase, a method of  
CC producing active heparanase and a method of modulating an adhesion  
CC activity of heparanase. The composition and methods are useful for  
CC modulating heparanase activation and for treating heparanase-associated  
CC diseases or disorders such as cancer, inflammation, cardiovascular  
CC diseases, neurological diseases or viral infections. This sequence  
CC represents a human heparanase protein used in the scope of the invention.  
XX  
XX Sequence 543 AA;  
SQ  
Query Match 100.0%; Score 82; DB 9; Length 543;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 SWELGNPNFLKKA 15  
DB 219 SWELGNPNFLKKA 233  
Search completed: June 5, 2006, 12:42:03  
Job time : 136.384 secs

GenCore version 5.1.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:43:17 ; Search time 16.6438 Seconds

(without alignments)  
86.714 Million cell updates/sec

Title: US-10-645-659A-8

Perfect score: 82

Sequence: 1 SWELGNPNFLKKA 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

1: Pir1.\*

2: Pir2.\*

3: Pir3.\*

4: Pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	54.9	648	2 A32576	beta-glucuronidase
2	43	52.4	624	2 T49366	myocyte-specific e
3	42	51.2	114	1 CCRFP7	cytochrome c2 - Rh
4	42	51.2	848	2 S48273	probable transcript
5	41.5	50.6	352	2 T08469	endo-1,4-beta-xyla
6	41	50.0	194	2 B83007	conserved hypothe
7	41	50.0	274	2 T19330	hypothetical prote
8	41	50.0	550	2 B84900	hypothetical prote
9	41	50.0	648	2 A25047	beta-glucuronidase
10	40.5	49.4	79	1 C36955	flagellar biosynth
11	40	48.8	117	2 B97391	hypothetical prote
12	40	48.8	117	2 AC2609	hypothetical prote
13	40	48.8	270	2 T32412	hypothetical prote
14	40	48.8	331	2 S54263	rep A protein - Ba
15	40	48.8	413	2 B27539	variant surface gl
16	40	48.8	505	2 B47059	cinnamate-4-hydrox
17	40	48.8	640	2 F81191	hypothetical prote
18	40	48.8	651	2 A26581	beta-glucuronidase
19	40	48.8	716	1 A35269	translation initia
20	40	48.8	1641	2 D82704	conserved hypothe
21	39	47.6	94	2 S48138	ubiquinol-cytochro
22	39	47.6	263	2 T22536	hypothetical prote
23	39	47.6	325	2 A84326	hypothetical prote
24	39	47.6	366	2 T04763	chitinase homolog
25	39	47.6	442	2 F87348	mannanase, probabl
26	39	47.6	459	2 F70189	rod shape-determin
27	39	47.6	473	1 T40391	phosphoprotein pho
28	39	47.6	481	2 H84619	hypothetical prote
29	39	47.6	545	2 S67621	hypothetical prote

30	39	47.6	554	2 H95922	hypothetical membr
31	39	47.6	628	2 AD2408	hypothetical prote
32	39	47.6	651	2 AD3057	glycogen debranchi
33	39	47.6	651	2 B98229	glycogen debranchi
34	39	47.6	660	2 I51684	epithelial sodium
35	39	47.6	669	2 D22278	endo-1,4-beta-mann
36	39	47.6	941	2 S29043	cellulase (EC 3.2.
37	38.5	47.0	642	2 E72215	oligopeptide ABC t
38	38	46.3	100	2 G96948	uncharacterized sm
39	38	46.3	129	2 H87375	hypothetical prote
40	38	46.3	135	2 F86239	protein F20824.4 [
41	38	46.3	139	2 AH1586	bacteriophage phi-
42	38	46.3	217	2 F82788	thymidylate kinase
43	38	46.3	228	2 AE2956	haloacid dehalogen
44	38	46.3	228	2 A98327	hypothetical prote
45	38	46.3	230	2 S49780	hypothetical prote
46	38	46.3	247	2 F70710	probable 3-oxoacyl
47	38	46.3	277	2 T38320	hypothetical prote
48	38	46.3	322	2 T24356	hypothetical prote
49	38	46.3	344	2 S09283	fructose-bisphosph
50	38	46.3	379	2 T04762	chitinase homolog
51	38	46.3	398	2 T04761	chitinase homolog
52	38	46.3	411	2 D86153	hypothetical prote
53	38	46.3	415	2 AG2044	hypothetical prote
54	38	46.3	464	2 T28662	hypothetical prote
55	38	46.3	581	2 B86318	protein F15H18.9 [
56	38	46.3	600	2 AE2570	hypothetical prote
57	38	46.3	661	2 S75005	sensory transducti
58	38	46.3	726	2 S37964	beta-adaptin homol
59	38	46.3	843	2 F96537	hypothetical prote
60	38	46.3	1140	2 S73786	hypothetical prote
61	38	46.3	1540	2 T45619	hypothetical prote
62	38	46.3	2828	2 S59413	hypothetical prote
63	38	46.3	3351	2 T13812	probable membrane
64	37.5	45.7	501	2 A21159	lipophorin - fruit
65	37	45.1	122	2 H82231	cell surface anti
66	37	45.1	128	2 S75259	hypothetical prote
67	37	45.1	143	2 B95408	probable DNA-bind
68	37	45.1	146	2 F90141	hypothetical prote
69	37	45.1	232	2 A27503	testis-specific pr
70	37	45.1	250	2 B40626	probable 2,3-dihyd
71	37	45.1	266	2 T05471	hypothetical prote
72	37	45.1	275	2 T40345	probable autophago
73	37	45.1	278	2 AF3430	ribonuclease III [
74	37	45.1	283	2 T50431	CBP3-like protein
75	37	45.1	308	2 T03862	hypothetical prote
76	37	45.1	310	2 S60425	probable membrane
77	37	45.1	342	2 AD2306	hypothetical prote
78	37	45.1	345	2 F86944	probable fructose
79	37	45.1	353	2 AB2483	hypothetical prote
80	37	45.1	363	2 A81016	transcription regu
81	37	45.1	395	2 G84259	histidinol-phospha
82	37	45.1	430	2 S61118	DIN7 protein - Yea
83	37	45.1	437	2 H70918	probable PE protei
84	37	45.1	442	2 F83367	probable glycosyl
85	37	45.1	446	2 T27696	hypothetical prote
86	37	45.1	466	2 S39494	glutathione-disulf
87	37	45.1	472	2 T51429	hypothetical prote
88	37	45.1	477	2 H86466	protein F23M19.7 [
89	37	45.1	499	2 T03213	probable ribosomal
90	37	45.1	521	2 T45608	hypothetical prote
91	37	45.1	538	2 T03232	probable ribosomal
92	37	45.1	541	2 G90516	ABC transporter at
93	37	45.1	564	2 F82194	hypothetical prote
94	37	45.1	631	2 S70910	transferrin-bindin
95	37	45.1	654	2 F83260	hypothetical prote
96	37	45.1	655	2 H82938	probable ABC subst
97	37	45.1	780	2 H84453	probable heat choc
98	37	45.1	781	2 S49340	heat-shock protein
99	37	45.1	804	2 B89961	leucyl-tRNA synthe
100	37	45.1	819	2 S15169	ferric-pseudobacti
101	37	45.1	844	2 C28667	DNA mismatch repai
102	37	45.1	857	2 E98107	DNA mismatch repai

103	37	45.1	1075	2	T47603	beta Galactosidase	176	35	42.7	251	2	C71879	hypothetical prote
104	37	45.1	1150	2	T15277	hypothetical prote	177	35	42.7	254	2	A80940	probable ribulose-
105	37	45.1	1661	2	H71439	hypothetical prote	178	35	42.7	255	2	B70982	probable nei prote
106	36.5	44.5	458	2	F98271	mraA protein [impo	179	35	42.7	261	2	F90506	hypothetical prote
107	36.5	44.5	458	2	A13012	phosphoglucosutase	180	35	42.7	272	2	A42168	hypothetical prote
108	36.5	44.5	475	2	AB3235	phosphomannosutase	181	35	42.7	282	2	I50996	connexin 32.7 - At
109	36	43.9	35	2	E95098	hypothetical prote	182	35	42.7	294	2	T00104	probable dTDP-4-de
110	36	43.9	138	2	S74872	hypothetical prote	183	35	42.7	298	2	B35272	osteoinductive fac
111	36	43.9	138	2	S77514	hypothetical prote	184	35	42.7	298	2	Jc4130	osteoinductive fac
112	36	43.9	138	2	S75128	hypothetical prote	185	35	42.7	299	2	A35272	osteoinductive fac
113	36	43.9	154	1	JC1292	protein-export pro	186	35	42.7	314	2	T28879	hypothetical prote
114	36	43.9	173	2	N0515	fliv protein - Sal	187	35	42.7	328	2	T24006	hypothetical prote
115	36	43.9	203	2	H90287	hypothetical prote	188	35	42.7	331	2	T35809	hypothetical prote
116	36	43.9	209	2	D90788	probable chaperone	189	35	42.7	341	2	T22057	hypothetical prote
117	36	43.9	229	2	E85648	probable chaperone	190	35	42.7	341	2	T26920	hypothetical prote
118	36	43.9	235	2	H71615	probable membrane	191	35	42.7	348	2	E87466	transcription regu
119	36	43.9	259	2	T19091	hypothetical prote	192	35	42.7	365	2	B33829	glycine oxidase (s
120	36	43.9	272	2	T48567	ABA-responsive pro	193	35	42.7	367	2	T36278	hypothetical prote
121	36	43.9	281	2	F84313	hypothetical prote	194	35	42.7	375	2	T11020	transaldolase - My
122	36	43.9	288	2	G64146	yabH protein homol	195	35	42.7	379	2	G70918	hypothetical prote
123	36	43.9	315	2	AF1393	glycosyl transfera	196	35	42.7	392	2	B69522	conserved hypotet
124	36	43.9	315	2	AF1168	glycosyl transfera	197	35	42.7	395	2	H69147	LPS biosynthesis R
125	36	43.9	326	2	S54267	repa protein - Bac	198	35	42.7	395	2	A86458	probable elicitor
126	36	43.9	335	2	T19662	hypothetical prote	199	35	42.7	398	2	B86353	protein F282.6 [im
127	36	43.9	362	2	D82644	sugar-phosphate de	200	35	42.7	399	2	H86453	avr9 homolog F9L11
128	36	43.9	369	2	T04323	mannan endo-1,4-be	201	35	42.7	403	2	A84592	(1-4)-beta-mannan
129	36	43.9	373	2	C70917	probable tal prote	202	35	42.7	406	2	T04817	hypothetical prote
130	36	43.9	414	2	S59783	hypothetical prote	203	35	42.7	425	2	E71075	probable proton gl
131	36	43.9	416	2	A97093	probable membrane	204	35	42.7	431	2	T09048	probable mannan en
132	36	43.9	418	2	S55018	CEN-1 protein - Ca	205	35	42.7	432	2	S51901	tryptophan-trNA 11
133	36	43.9	422	2	S49012	yolk protein 3 - b	206	35	42.7	432	2	C82146	periplasmic sugar-
134	36	43.9	448	2	T48214	endo-1,4-beta-mann	207	35	42.7	440	2	JC8032	PU.1-binding prote
135	36	43.9	484	2	AF3107	xylokkinase [impo	208	35	42.7	444	2	C81054	phosphoglucosutase
136	36	43.9	488	2	F98179	D-arabinitol kinas	209	35	42.7	445	2	T39415	probable transcrip
137	36	43.9	519	2	T46241	hypothetical prote	210	35	42.7	446	2	H86899	Glutamate-ammonia
138	36	43.9	525	2	AF3601	periplasmic oligop	211	35	42.7	448	2	D97927	glutamate-ammonia
139	36	43.9	528	2	D87443	alkaline phosphata	212	35	42.7	448	2	C95058	glutamine syntheta
140	36	43.9	552	1	K7HUGU	N-acetylglucosamin	213	35	42.7	455	2	S16559	cellulase (EC 3.2.
141	36	43.9	554	2	T50118	kinesin-related pr	214	35	42.7	458	2	S75114	hypothetical prote
142	36	43.9	629	2	S31174	hypothetical prote	215	35	42.7	465	2	H69120	hypothetical prote
143	36	43.9	662	2	B72114	DNA ligase CP0624	216	35	42.7	478	2	C86274	hypothetical prote
144	36	43.9	662	2	E86509	DNA ligase [impor	217	35	42.7	484	2	T34238	hypothetical prote
145	36	43.9	741	2	A27892	translation initia	218	35	42.7	504	2	G82631	glutamine syntheta
146	36	43.9	746	2	C95110	competence protein	219	35	42.7	509	1	S62899	cytochrome P450 (C
147	36	43.9	746	2	A97979	competence protein	220	35	42.7	511	1	VGBE1K	glycoprotein C - h
148	36	43.9	807	2	E90523	leucyl-trna synthe	221	35	42.7	511	1	VGBEF4	hypothetical prote
149	36	43.9	899	2	G36812	hypothetical prote	222	35	42.7	512	2	T13611	hypothetical prote
150	36	43.9	931	2	G69615	ATP-dependent DNA	223	35	42.7	512	2	AC3203	IS3 family transpo
151	36	43.9	989	2	T47503	hypothetical prote	224	35	42.7	512	2	AD2835	IS3 family transpo
152	36	43.9	1084	2	H86229	hypothetical prote	225	35	42.7	512	2	AD3049	IS3 family transpo
153	36	43.9	1124	1	FKPUZ	phytochrome - zucc	226	35	42.7	512	2	G97612	probable transposa
154	36	43.9	1169	2	G72571	probable DNA-direc	227	35	42.7	512	2	G98236	probable transposa
155	36	43.9	1184	2	H86190	hypothetical prote	228	35	42.7	548	2	AC0871	probable integrase
156	36	43.9	1274	2	T04018	hypothetical prote	229	35	42.7	551	2	H86260	type III secretion
157	36	43.9	1832	2	C82594	glutamate synthase	230	35	42.7	567	2	E91095	type III secretion
158	36	43.9	1858	2	D97376	hypothetical prote	231	35	42.7	567	2	A85941	hypothetical prote
159	36	43.9	1939	2	AF0095	probable sideropho	232	35	42.7	584	2	S40013	glucan 1,4 alpha g
160	35.5	43.3	217	2	B83334	reverse cytochrom	233	35	42.7	622	2	G90250	probable formyl tr
161	35.5	43.3	470	2	E97331	Hfl1 protein - yea	234	35	42.7	667	2	AD0295	probable formyl tr
162	35.5	43.3	488	2	S61010	hypothetical prote	235	35	42.7	668	2	A44863	UDPglucose 4-epime
163	35.5	43.3	726	2	T13348	hypothetical prote	236	35	42.7	689	1	S29621	probable chitinase
164	35	42.7	108	2	T17630	hypothetical prote	237	35	42.7	699	2	AC0504	hypothetical prote
165	35	42.7	154	2	AB1605	B. subtilis comG o	238	35	42.7	699	2	T15920	probable membrane
166	35	42.7	176	2	A56652	interleukin-7 prec	239	35	42.7	706	2	S61717	NADH2 dehydrogenas
167	35	42.7	176	2	JC4035	interleukin 7 prot	240	35	42.7	732	2	T14233	NADH2 dehydrogenas
168	35	42.7	177	2	C95922	hypothetical prote	241	35	42.7	732	2	T13814	NADH2 dehydrogenas
169	35	42.7	208	2	T12985	hypothetical prote	242	35	42.7	737	2	T14232	topoisomerase iv c
170	35	42.7	217	2	S62795	probable lipoprote	243	35	42.7	750	2	A97501	topoisomerase iv s
171	35	42.7	219	2	H97110	phosphate uptake r	244	35	42.7	750	2	BE2719	zinc finger protei
172	35	42.7	230	2	D71601	probable secreted r	245	35	42.7	805	2	B35047	protein kinase (EC
173	35	42.7	247	2	AC2786	SEC-independent pr	246	35	42.7	830	2	B44439	adenylyl cyclase t
174	35	42.7	247	2	E97565	hypothetical prote	247	35	42.7	839	2	A47007	
175	35	42.7	251	2	F64636	conserved hypotet	248	35	42.7				

249	35	42.7	856	2	A44439	protein kinase (EC	322	34	41.5	248	2	T04758	hypothetical prote
250	35	42.7	857	2	A42861	protein kinase TTK	323	34	41.5	252	2	D64564	hypothetical prote
251	35	42.7	861	1	S34730	1,4-alpha-glucan b	324	34	41.5	252	2	B71947	hypothetical prote
252	35	42.7	1002	2	B91180	probable oxidoredu	325	34	41.5	261	2	H71144	hypothetical prote
253	35	42.7	1002	2	C86026	probable oxidoredu	326	34	41.5	269	2	T44866	hypothetical prote
254	35	42.7	1022	2	T24663	hypothetical prote	327	34	41.5	283	2	H70104	chemotaxis protein
255	35	42.7	1044	2	T43155	nitrite reductase	328	34	41.5	288	2	A75638	probable transposa
256	35	42.7	1044	2	T43160	nitrite reductase	329	34	41.5	288	2	C75426	probable transposa
257	35	42.7	1087	2	T16876	hypothetical prote	330	34	41.5	293	2	T32892	hypothetical prote
258	35	42.7	1115	2	T19137	hypothetical prote	331	34	41.5	297	2	S65185	ribosomal protein
259	35	42.7	1121	2	T138127	phosphoprotein pho	332	34	41.5	304	2	T15729	hypothetical prote
260	35	42.7	1122	1	FKMUA	phytochrome A - Ar	333	34	41.5	305	2	B33282	DNA-binding protei
261	35	42.7	1122	2	D86229	phytochrome A [imp	334	34	41.5	311	2	D69229	conserved hypotet
262	35	42.7	1134	2	A41350	adenylate cyclase	335	34	41.5	319	2	T42423	PAS7 protein homol
263	35	42.7	1166	1	E65151	hypothetical 126K	336	34	41.5	322	2	AD3488	transporter BME18
264	35	42.7	1232	2	T43027	neural cell adhesi	337	34	41.5	327	2	E75618	probable transposa
265	35	42.7	1244	2	A05218	hypothetical prote	338	34	41.5	327	2	C75556	probable transposa
266	35	42.7	1259	2	S25954	gene atpA intron 2	339	34	41.5	327	2	C75624	probable transposa
267	35	42.7	1274	2	T16251	hypothetical prote	340	34	41.5	327	2	B75620	probable transposa
268	35	42.7	1305	2	H41662	150K mating aggreg	341	34	41.5	327	2	A75631	probable transposa
269	35	42.7	1324	2	S51622	cut3 protein - fis	342	34	41.5	327	2	A75633	probable transposa
270	35	42.7	1480	2	T21911	hypothetical prote	343	34	41.5	336	2	T26993	hypothetical prote
271	35	42.7	1483	2	T21914	hypothetical prote	344	34	41.5	340	2	T32891	hypothetical prote
272	35	42.7	1483	2	T21912	hypothetical prote	345	34	41.5	344	2	D70576	probable fructose
273	35	42.7	1510	2	C84727	probable glucan sy	346	34	41.5	344	2	H70030	conserved hypotet
274	35	42.7	1729	2	T18396	erythrocyte membra	347	34	41.5	350	2	A95383	probable ABC trans
275	35	42.7	1788	2	E86189	hypothetical prote	348	34	41.5	354	2	S15660	(2'-5')oligo(A) sy
276	35	42.7	3191	2	T22945	hypothetical prote	349	34	41.5	358	2	S44233	strp protein - str
277	35	42.7	3300	2	D70575	probable ppg prote	350	34	41.5	366	2	T34662	probable integral
278	34.5	42.1	59	2	T26375	hypothetical prote	351	34	41.5	368	2	T47894	actin-like protein
279	34.5	42.1	115	2	T07576	hypothetical prote	352	34	41.5	375	2	JX0131	cellulase (EC 3.2.
280	34.5	42.1	150	2	T10641	carbonic anhydrase	353	34	41.5	383	2	T72352	hypothetical prote
281	34.5	42.1	201	1	A60598	smooth muscle prot	354	34	41.5	385	2	G69101	mannose-1-phosphat
282	34.5	42.1	201	2	A57015	Gene SM22 alpha pr	355	34	41.5	387	2	T22511	hypothetical prote
283	34.5	42.1	372	2	G96973	probable membrane	356	34	41.5	387	2	A86456	unknown protein [i
284	34.5	42.1	398	2	C81729	Mtr/TnaB/TyO perm	357	34	41.5	398	2	JQ0366	phospholipase C (E
285	34.5	42.1	415	2	C71467	probable tyrosine	358	34	41.5	398	2	B49231	phospholipase C, a
286	34.5	42.1	596	2	S32802	apolipoprotein B -	359	34	41.5	398	2	A30565	pectin acetyl ester
287	34.5	42.1	904	2	S51299	hypothetical prote	360	34	41.5	422	2	T01197	hypothetical prote
288	34.5	42.1	1448	2	T08526	DNA primase traC2	361	34	41.5	426	1	T49035	acid phosphatase (
289	34.5	42.1	1448	2	S37669	traC-2 protein - E	362	34	41.5	443	2	AB1405	probable phosphoes
290	34.5	42.1	1687	2	T39072	DNA2-NAM7 helicase	363	34	41.5	443	2	AI1780	probable phosphoes
291	34.5	42.1	2054	2	T07584	hypothetical prote	364	34	41.5	448	2	T51400	hypothetical prote
292	34	41.5	40	2	A29184	vitellogenin - tur	365	34	41.5	474	2	T26990	hypothetical prote
293	34	41.5	116	2	C59100	hypothetical prote	366	34	41.5	475	2	AI0010	probable membrane
294	34	41.5	122	2	S32574	serum amyloid prot	367	34	41.5	477	2	T01202	probable RNA helic
295	34	41.5	132	2	A44864	fatty acid-binding	368	34	41.5	504	2	T52171	cytochrome P450 mo
296	34	41.5	134	2	E83549	hypothetical prote	369	34	41.5	505	2	G75197	lacZ expression re
297	34	41.5	138	2	F75056	hypothetical prote	370	34	41.5	505	2	F71230	hypothetical prote
298	34	41.5	138	2	E71175	hypothetical prote	371	34	41.5	507	2	E90539	hypothetical prote
299	34	41.5	154	2	H89975	hypothetical prote	372	34	41.5	519	2	S45723	P60 protein - oat
300	34	41.5	156	1	A55446	protein-tyrosine-p	373	34	41.5	530	2	JC7979	cellubiohydrolase
301	34	41.5	159	2	H88003	hypothetical prote	374	34	41.5	563	2	A72300	beta-glucuronidase
302	34	41.5	161	2	F83297	protein KllB4.2 [i	375	34	41.5	573	2	T49315	receptor lectin ki
303	34	41.5	162	2	S73299	phycocyanin alpha	376	34	41.5	574	2	S50756	beta-D-glucosidase
304	34	41.5	170	2	H96931	tryptophan-rich pr	377	34	41.5	615	2	S38088	hypothetical prote
305	34	41.5	194	2	S75290	hypothetical prote	378	34	41.5	629	2	T47798	receptor lectin ki
306	34	41.5	194	2	S76919	hypothetical prote	379	34	41.5	636	2	S47299	gene F protein - r
307	34	41.5	200	1	A26694	smooth muscle prot	380	34	41.5	631	2	FC1123	hypothetical prote
308	34	41.5	201	1	JN0774	smooth muscle prot	381	34	41.5	653	2	T34356	hypothetical prote
309	34	41.5	201	1	JS0774	smooth muscle prot	382	34	41.5	656	2	S76505	hypothetical prote
310	34	41.5	210	2	T03865	hypothetical prote	383	34	41.5	666	2	T43171	cytoplasmic signal
311	34	41.5	211	2	A80935	probable TetR-fam	384	34	41.5	667	2	E86728	NADH dehydrogenase
312	34	41.5	215	2	AG0476	Tetr-family transc	385	34	41.5	672	2	T24507	hypothetical prote
313	34	41.5	223	2	AH0890	disulfide isomeras	386	34	41.5	692	2	B70484	conserved hypotet
314	34	41.5	224	2	AH1815	two-component resp	387	34	41.5	748	2	T51738	RNA helicase RH3 [
315	34	41.5	233	1	F69178	conserved hypotet	388	34	41.5	749	2	G86186	hypothetical prote
316	34	41.5	234	2	F65203	hypothetical 26.6	389	34	41.5	783	2	JC5467	cellulase (EC 3.2.
317	34	41.5	234	2	F91240	hypothetical prote	390	34	41.5	785	2	G95825	hypothetical prote
318	34	41.5	234	2	C86088	hypothetical prote	391	34	41.5	785	2	T52059	ent-kaurane syntha
319	34	41.5	239	2	AC2745	glycerophosphoryl	392	34	41.5	800	2	A29003	cellulase (EC 3.2.
320	34	41.5	246	2	T37473	transcription regu	393	34	41.5	800	2	T23780	hypothetical prote
321	34	41.5	246	2	B97526	hypothetical prote	394	34	41.5	822	2	T41941	glycoprotein B - h

395	34	41.5	822	2	JT0611	cellulase (EC 3.2.	468	33	40.2	268	2	AG2603	components of type
396	34	41.5	825	2	JC7532	cellulase (EC 3.2.	469	33	40.2	268	2	AG7385	cpab protein (Af22
397	34	41.5	824	2	JS0174	cellulase (EC 3.2.	470	33	40.2	270	2	AF2170	inositol monophosph
398	34	41.5	839	2	T21207	hypothetical prote	471	33	40.2	272	2	G71115	hypothetical prote
399	34	41.5	866	2	I79267	traB protein - Bsc	472	33	40.2	275	2	G87125	probable TetR-fami
400	34	41.5	888	2	E82885	hypothetical prote	473	33	40.2	276	2	AD1765	hypothetical cell
401	34	41.5	898	2	H84701	probable villin li	474	33	40.2	277	2	AD1390	hypothetical cell
402	34	41.5	899	2	T42976	hypothetical prote	475	33	40.2	283	2	E90571	conserved hypothet
403	34	41.5	1097	2	T45622	hypothetical prote	476	33	40.2	287	2	S71279	L-ascorbate peroxi
404	34	41.5	1184	1	A34795	kinesin-related pr	477	33	40.2	288	2	S60438	protein N-acetyltr
405	34	41.5	1216	2	F88473	protein F40H6.5 li	478	33	40.2	291	2	AH1680	methytransferases
406	34	41.5	1229	2	H84465	hypothetical prote	479	33	40.2	292	2	E88482	protein COSD11.9 l
407	34	41.5	1239	2	T42020	class IV chitin sy	480	33	40.2	296	2	I40818	phosphotransbutyry
408	34	41.5	1250	2	T40062	probable nuclear e	481	33	40.2	301	2	E97278	phosphate butyrylt
409	34	41.5	1253	2	T14349	Shyc protein - mou	482	33	40.2	310	2	G02309	UDP-glucuronosyltr
410	34	41.5	1323	2	T11661	phosphoribosylform	483	33	40.2	311	2	I64082	pantothenate kinas
411	34	41.5	1339	2	JC5508	DNA-directed DNA p	484	33	40.2	311	2	F98245	n-formylglutamate
412	34	41.5	1451	2	S42167	190K protein - hum	485	33	40.2	312	2	T02406	hypothetical prote
413	34	41.5	1666	2	A48594	skelemin - mouse	486	33	40.2	314	2	A99446	ABC transporter, A
414	34	41.5	1876	2	T13801	phosphoinositide 3	487	33	40.2	317	2	JC5696	prolyl aminopeptid
415	34	41.5	2358	2	T39569	probable alpha-glu	488	33	40.2	319	2	T38533	sur1 protein homol
416	34	41.5	2371	2	T43432	alpha-glucan synth	489	33	40.2	330	1	JN0561	urkinase-type pla
417	34	41.5	3573	2	S23070	erythronolide synt	490	33	40.2	330	2	D95362	probable AraC-fami
418	34	41.5	5032	1	A35041	ryanodine receptor	491	33	40.2	337	1	E70191	conserved hypothet
419	34	41.5	5037	2	B35041	microtubule-associ	492	33	40.2	340	2	T26923	hypothetical prote
420	34	41.5	5327	2	T13564	hypothetical prote	493	33	40.2	348	1	S32521	alcohol dehydrogen
421	33.5	40.9	85	2	JQ0278	hypothetical prote	494	33	40.2	348	2	C83571	probable binding p
422	33.5	40.9	85	2	S58627	hypothetical prote	495	33	40.2	351	2	T26918	hypothetical prote
423	33.5	40.9	88	2	T43580	type III secretion	496	33	40.2	353	2	AF2560	hypothetical prote
424	33.5	40.9	146	2	S54789	superoxide dismuta	497	33	40.2	355	2	A12867	endo-1,4-beta-xyla
425	33.5	40.9	159	2	A90546	hypothetical prote	498	33	40.2	357	2	F82878	XAA-PRO aminopept
426	33.5	40.9	174	2	B75114	probable NADH dehy	499	33	40.2	358	2	H36891	transfer complex p
427	33.5	40.9	175	2	T14999	hypothetical prote	500	33	40.2	359	2	E72378	sugar ABC transpor
428	33.5	40.9	259	2	T20399	hypothetical prote	501	33	40.2	365	2	E97644	endo-1,4-beta-xyla
429	33.5	40.9	289	2	AD2519	hypothetical prote	502	33	40.2	368	2	A85768	partial beta-D-glu
430	33.5	40.9	332	2	G90291	endoglucanase prec	503	33	40.2	370	2	D90919	beta-D-glucuronida
431	33.5	40.9	334	2	G90360	conserved hypothet	504	33	40.2	376	2	G84404	hypothetical prote
432	33.5	40.9	486	2	A69442	conserved hypothet	505	33	40.2	377	2	T06192	probable endo-1,4-
433	33.5	40.9	648	2	T47895	hypothetical prote	506	33	40.2	379	2	A49679	exo-alpha-sialidas
434	33.5	40.9	837	2	A34898	granulocyte colony	507	33	40.2	381	2	H75173	udp-n-acetylglucos
435	33.5	40.9	891	2	AC0149	DNA topoisomerase	508	33	40.2	381	2	C96657	hypothetical prote
436	33.5	40.9	940	2	T00056	hypothetical prote	509	33	40.2	388	1	WMLJBT	membrane-bound lyt
437	33.5	40.9	1662	2	T18540	moA protein precu	510	33	40.2	390	2	AH0125	hypothetical prote
438	33	40.2	35	2	T07870	major latex protei	511	33	40.2	394	2	A86431	nuclear protein EM
439	33	40.2	51	2	S78688	hypothetical prote	512	33	40.2	397	2	T09579	prescorin-5Y methy
440	33	40.2	109	2	T00234	hypothetical prote	513	33	40.2	398	2	T44688	alkane 1-monooxyge
441	33	40.2	114	1	CCRF2P	cytochrome c2 - Rh	514	33	40.2	401	1	A31266	Trp-Asp repeat pro
442	33	40.2	124	2	B47098	binding-protein-de	515	33	40.2	404	2	T40553	DNA polymerase pro
443	33	40.2	147	2	A70462	ribosomal protein	516	33	40.2	406	2	T42561	cellulase (EC 3.2.
444	33	40.2	151	2	H64104	dupP diphosphatase	517	33	40.2	409	2	B25156	cellulase (EC 3.2.
445	33	40.2	169	2	H83972	hypothetical prote	518	33	40.2	427	2	A41790	actin like protein
446	33	40.2	173	2	G81428	hypothetical prote	519	33	40.2	427	2	T06191	probable endo-1,4-
447	33	40.2	177	2	C71329	hypothetical prote	520	33	40.2	427	2	T06195	probable endo-1,4-
448	33	40.2	185	2	T36546	hypothetical prote	521	33	40.2	429	2	T14417	S-locus-specific g
449	33	40.2	189	2	S41410	insecticynanin prec	522	33	40.2	430	2	T14420	S-locus-specific g
450	33	40.2	190	2	H71245	hypothetical prote	523	33	40.2	441	2	AB1290	ATP-dependent RNA
451	33	40.2	196	2	A39021	pinin - Haemophilu	524	33	40.2	442	2	AH1661	ATP-dependent RNA
452	33	40.2	211	2	T38645	hypothetical prote	525	33	40.2	451	2	F87039	protoporphyrinogen
453	33	40.2	211	2	G90209	conserved hypothet	526	33	40.2	463	2	T16218	translation elonga
454	33	40.2	216	2	A60331	pinin precursor -	527	33	40.2	466	2	C42360	cellulase (EC 3.2.
455	33	40.2	222	2	C69326	conserved hypothet	528	33	40.2	471	1	A35867	cytochrome P450 71
456	33	40.2	227	2	A48412	hypothetical prote	529	33	40.2	471	2	T47568	fructokinase-like
457	33	40.2	236	2	I67432	BCL-2 - rat (fragm	530	33	40.2	473	2	S76985	hypothetical prote
458	33	40.2	236	2	A90190	hypothetical prote	531	33	40.2	474	2	S56748	glutathione syntha
459	33	40.2	237	2	T05249	hypothetical prote	532	33	40.2	478	2	C83739	hypothetical prote
460	33	40.2	238	2	AD2034	hypothetical prote	533	33	40.2	486	2	I40548	bifunctional cellu
461	33	40.2	245	2	E70861	hypothetical prote	534	33	40.2	488	2	F85763	amino acid permeas
462	33	40.2	253	1	VHVUTV	nucleocapsid prote	535	33	40.2	488	2	A25156	cellulase (EC 3.2.
463	33	40.2	253	2	S47480	chlorophyll a/b-bi	536	33	40.2	489	1	A53766	phosphoinositide-s
464	33	40.2	253	2	T04642	hypothetical prote	537	33	40.2	494	2	G64382	hypothetical synth
465	33	40.2	254	2	H97081	proline/glycine be	538	33	40.2	494	2	T04961	acetylactate synth
466	33	40.2	255	2	S52341	LHCI-680, photosys	539	33	40.2	498	2	T50525	cytochrome P450 mo
467	33	40.2	266	2	AF3040	conserved hypothet	540	33	40.2	499	2	JN0111	cellulase (EC 3.2.

541	33	40.2	499	2	A27198	cellulase (EC 3.2.	614	33	40.2	992	2	T39315	hypothetical prote
542	33	40.2	501	2	G86460	probable cytochrom	615	33	40.2	1050	2	S54640	KCS1 protein - yea
543	33	40.2	502	1	S45039	cytochrome P450-	616	33	40.2	1053	2	T07965	reverse transcript
544	33	40.2	502	2	T52256	cytochrome P-450LX	617	33	40.2	1057	2	T16676	hypothetical prote
545	33	40.2	502	2	T07141	cytochrome P450 CY	618	33	40.2	1063	2	A53164	glutamate dehydrog
546	33	40.2	504	2	H83791	propionyl-CoA carb	619	33	40.2	1103	2	T13590	distal tail fiber
547	33	40.2	504	2	D86332	hypothetical prote	620	33	40.2	1130	2	A48843	MHC class II trans
548	33	40.2	504	2	S54744	cellulase (EC 3.2.	621	33	40.2	1238	2	T32625	hypothetical prote
549	33	40.2	505	2	S39962	endoglucanase - Er	622	33	40.2	1243	2	S07278	tail fiber protein
550	33	40.2	508	2	A1453	hypothetical prote	623	33	40.2	1247	2	T45743	hypothetical prote
551	33	40.2	508	2	G69593	cellulase (EC 3.2.	624	33	40.2	1252	2	D71810	probable type II D
552	33	40.2	508	2	A36874	cellulase (EC 3.2.	625	33	40.2	1294	2	T04278	hypothetical prote
553	33	40.2	510	2	T07119	cytochrome P450 Cp	626	33	40.2	1342	1	S32680	DNA-directed RNA p
554	33	40.2	512	2	T06065	probable cytochrom	627	33	40.2	1364	1	AJF8FP	phosphoribosylamin
555	33	40.2	512	2	T00870	probable cytochrom	628	33	40.2	1465	2	T30891	PH3 protein - mai
556	33	40.2	516	2	D4563	single-stranded-DN	629	33	40.2	1590	2	B87754	protein C43E11.3 (
557	33	40.2	516	2	B71946	probable single-str	630	33	40.2	1601	2	AE2011	hypothetical prote
558	33	40.2	532	2	T10624	reticuline oxidase	631	33	40.2	1622	2	D86428	glutathione S-conj
559	33	40.2	545	2	A10265	periplasmic oligop	632	33	40.2	1623	2	T01369	ABC transporter At
560	33	40.2	547	2	JQ0356	cellulase (EC 3.2.	633	33	40.2	1631	1	SAZQK1	major merozoite su
561	33	40.2	547	2	T44743	probable thiamin b	634	33	40.2	1639	2	S05603	major merozoite su
562	33	40.2	548	2	S33788	Photinus-luciferin	635	33	40.2	1640	2	A24594	probable major sur
563	33	40.2	560	2	C84632	hypothetical prote	636	33	40.2	1662	2	T01893	hypothetical prote
564	33	40.2	576	2	T48585	auxin-regulated pr	637	33	40.2	1674	2	G96736	hypothetical prote
565	33	40.2	579	2	JT0494	alpha-glucosidase	638	33	40.2	1726	1	SAZQGM	major merozoite su
566	33	40.2	583	2	T25690	hypothetical prote	639	33	40.2	1726	2	A45948	major merozoite su
567	33	40.2	594	2	C70021	butyryl-CoA dehydr	640	33	40.2	1739	2	A48298	sodium channel hom
568	33	40.2	600	2	E53290	oligopeptide trans	641	33	40.2	1786	1	MMMSB1	laminin beta-1 cha
569	33	40.2	603	1	G8EC9C	beta-glucuronidase	642	33	40.2	1993	2	AF1450	probable peptidogl
570	33	40.2	606	2	H97012	hypothetical prote	643	33	40.2	2004	2	D88948	protein ZK1005.1 (
571	33	40.2	620	2	T30765	hypothetical prote	644	33	40.2	2437	2	S53611	MIBP1 protein - ra
572	33	40.2	631	2	C90552	lipoprotein (impor	645	33	40.2	2480	2	D84904	hypothetical prote
573	33	40.2	636	2	H97148	molybdopterin bios	646	33	40.2	2697	2	T25444	hypothetical prote
574	33	40.2	640	2	G96034	conserved hypotet	647	33	40.2	3738	2	T05501	hypothetical prote
575	33	40.2	641	2	JC7331	gamma-glutamyltran	648	33	40.2	5005	2	F82884	hypothetical prote
576	33	40.2	649	2	T47609	hypothetical prote	649	32.5	39.6	151	2	H97133	phage terminase-li
577	33	40.2	652	2	A42245	thiamin biosynthes	650	32.5	39.6	201	2	T22600	hypothetical prote
578	33	40.2	659	2	A45184	protein-tyrosine k	651	32.5	39.6	235	2	T04451	hypothetical prote
579	33	40.2	659	2	I49553	protein-tyrosine k	652	32.5	39.6	239	2	S65825	hypothetical prote
580	33	40.2	665	2	C71667	propionyl-CoA carb	653	32.5	39.6	250	2	E75376	2-hydroxyhepta-2,4
581	33	40.2	665	2	G97819	hypothetical prote	654	32.5	39.6	254	2	H83334	probable transcrip
582	33	40.2	666	2	T35864	hypothetical prote	655	32.5	39.6	263	2	D86644	hypothetical prote
583	33	40.2	668	2	S55023	brown protein - fr	656	32.5	39.6	274	2	A72241	endoglucanase - Th
584	33	40.2	677	2	A32611	beta-galactosidase	657	32.5	39.6	282	2	G95130	transcription regu
585	33	40.2	684	2	T47694	probable serine/th	658	32.5	39.6	332	2	S27726	hypothetical prote
586	33	40.2	711	2	A37051	outer membrane pro	659	32.5	39.6	335	2	JT0569	chondromodulin-I p
587	33	40.2	719	2	JC1200	alpha-glucosidase	660	32.5	39.6	381	2	F81073	succinyl-diaminopi
588	33	40.2	720	2	S70838	piIQ protein - Nei	661	32.5	39.6	382	2	S32148	exo-alpha-sialidas
589	33	40.2	732	2	B84902	hypothetical prote	662	32.5	39.6	382	2	S01339	exo-alpha-sialidas
590	33	40.2	733	2	G87060	conserved membrane	663	32.5	39.6	450	2	F65183	4-alpha-l-fucosylt
591	33	40.2	739	2	H72364	aspartokinase II -	664	32.5	39.6	452	2	AB0921	probable 4-alpha-l
592	33	40.2	741	2	A26572	bsg25D protein - f	665	32.5	39.6	478	1	C42790	cystathionine beta
593	33	40.2	742	2	F84643	hypothetical prote	666	32.5	39.6	504	2	D71615	hypothetical prote
594	33	40.2	756	2	AB1088	chitinase B homolo	667	32.5	39.6	551	1	A55760	cystathionine beta
595	33	40.2	756	2	AB1452	chitinase B homolo	668	32.5	39.6	561	1	A42790	cystathionine beta
596	33	40.2	758	2	T51335	subtilisin-like pr	669	32.5	39.6	661	1	G71063	probable ferrous i
597	33	40.2	780	2	B70112	DNA mismatch repai	670	32.5	39.6	661	2	B75084	ferrous iron trans
598	33	40.2	782	2	A82940	hypothetical prote	671	32.5	39.6	770	1	TWBYA2	transcription fact
599	33	40.2	829	2	T19514	hypothetical prote	672	32.5	39.6	1220	2	A56136	jagged protein pre
600	33	40.2	860	2	T37768	probable vacuolar	673	32.5	39.6	1290	2	A56493	leucocyte common a
601	33	40.2	864	2	A81000	DNA mismatch repai	674	32.5	39.6	1898	2	S46216	hypothetical prote
602	33	40.2	864	2	C82019	DNA mismatch repai	675	32	39.0	68	2	G81063	hypothetical prote
603	33	40.2	876	1	ITKBAP	DNA topoisomerase	676	32	39.0	99	2	A32716	ig heavy chain C r
604	33	40.2	878	2	A41055	ecdysone receptor	677	32	39.0	102	2	S19464	hypothetical prote
605	33	40.2	880	1	S41420	valine-tRNA ligase	678	32	39.0	103	2	A64012	hypothetical prote
606	33	40.2	884	2	H83322	hypothetical prote	679	32	39.0	110	2	H75048	DNA-directed RNA p
607	33	40.2	932	2	F69552	leucyl-tRNA synthe	680	32	39.0	110	2	S50521	hypothetical prote
608	33	40.2	938	2	T01809	hypothetical prote	681	32	39.0	111	2	B32476	hypothetical prote
609	33	40.2	955	2	E84845	probable villin 2	682	32	39.0	122	2	B65570	L14 ribosomal prot
610	33	40.2	968	2	T01733	hypothetical prote	683	32	39.0	122	2	G72054	ribosomal protein
611	33	40.2	972	1	GNXSIV	genome polyprotein	684	32	39.0	126	2	AG8828	holo-lacyl-carrier
612	33	40.2	972	2	T09624	genome polypeptide	685	32	39.0	130	2	AH2424	hypothetical prote
613	33	40.2	976	2	T50669	villin 2 [imported	686	32	39.0	135	2	B72750	hypothetical prote



687	32	39.0	138	2	AH1214	hypothenical prote	760	32	39.0	315	2	AH1480	transcription regu
688	32	39.0	141	2	H70698	hypothenical prote	761	32	39.0	317	2	S51572	moCA protein - Rhl
689	32	39.0	141	2	QJ1669	hypothenical 16.1K	762	32	39.0	317	2	G72416	sugar ABC transpor
690	32	39.0	151	2	D81814	hypothenical prote	763	32	39.0	317	2	D95204	conserved hypothen
691	32	39.0	155	2	E64344	hypothenical prote	764	32	39.0	317	2	E98071	conserved hypothen
692	32	39.0	155	2	E69808	protein-tyrosine p	765	32	39.0	320	2	E71696	rare lipoprotein A
693	32	39.0	165	2	T46052	ADP-ribosylation f	766	32	39.0	324	2	B86819	oxidoeductase Ypj
694	32	39.0	171	2	T44537	hypothenical prote	767	32	39.0	325	2	D95845	conserved hypothen
695	32	39.0	173	2	AJ0432	transformation com	768	32	39.0	328	2	T20220	hypothenical prote
696	32	39.0	173	2	H70133	hypothenical prote	769	32	39.0	329	2	F97045	phage related prot
697	32	39.0	177	2	F71237	hypothenical prote	770	32	39.0	331	2	T01824	hypothenical prote
698	32	39.0	181	2	T08793	hypothenical prote	771	32	39.0	333	2	B47677	hypothenical prote
699	32	39.0	182	2	E81440	probable acetylra	772	32	39.0	334	2	C71718	hypothenical prote
700	32	39.0	187	2	T05608	hypothenical prote	773	32	39.0	336	2	G71900	hypothenical prote
701	32	39.0	190	2	T46053	ADP-ribosylation f	774	32	39.0	337	2	C71003	hypothenical prote
702	32	39.0	190	2	H85357	hypothenical prote	775	32	39.0	338	2	G71220	hypothenical prote
703	32	39.0	195	2	A82755	conserved hypothen	776	32	39.0	339	2	H75187	hypothenical prote
704	32	39.0	203	2	T02868	probable GTP-bind	777	32	39.0	340	2	F70852	virulence-regulati
705	32	39.0	205	2	S71584	GTP-binding protei	778	32	39.0	341	2	T40424	hypothenical prote
706	32	39.0	207	2	T17852	hypothenical prote	779	32	39.0	345	2	F97194	N-acetyl-gamma-glu
707	32	39.0	208	2	T04884	hypothenical prote	780	32	39.0	346	2	H72237	hypothenical prote
708	32	39.0	212	2	D71140	hypothenical prote	781	32	39.0	348	2	T29515	hypothenical prote
709	32	39.0	213	2	A41789	glutathione transf	782	32	39.0	348	2	C70415	cation efflux syst
710	32	39.0	214	2	D59108	hypothenical prote	783	32	39.0	350	2	A84421	ornithine transcar
711	32	39.0	215	2	F90312	conserved hypothen	784	32	39.0	352	2	JC5388	replication initia
712	32	39.0	216	2	T27851	hypothenical prote	785	32	39.0	355	2	A83358	ORF355 - Bradyrhiz
713	32	39.0	216	2	T29039	hypothenical prote	786	32	39.0	358	2	A83249	probable initiatio
714	32	39.0	217	2	H64186	hypothenical prote	787	32	39.0	359	2	T42087	probable 6-phospho
715	32	39.0	218	2	D97536	hypothenical prote	788	32	39.0	360	2	G86435	protein F17F8.7 [i
716	32	39.0	220	1	QVZC9	F9 protein - sheep	789	32	39.0	363	2	T26172	hypothenical prote
717	32	39.0	223	2	T46018	hypothenical prote	790	32	39.0	365	2	A75375	UDPgalactopyranose
718	32	39.0	230	2	T46018	hypothenical prote	791	32	39.0	367	1	I69653	probable cystathio
719	32	39.0	230	2	A86644	hypothenical prote	792	32	39.0	371	2	B71104	hypothenical prote
720	32	39.0	233	2	S72449	poly(beta-D-mannur	793	32	39.0	371	2	T02102	hypothenical prote
721	32	39.0	236	2	AG2905	hypothenical prote	794	32	39.0	377	2	JC7977	membrane-associate
722	32	39.0	236	2	H97680	probable ATP-bind	795	32	39.0	388	2	T21082	hypothenical prote
723	32	39.0	240	2	AC1269	uroporphyrinogen I	796	32	39.0	389	2	E86634	hypothenical prote
724	32	39.0	240	2	AE1631	uroporphyrinogen I	797	32	39.0	397	2	D91090	probable resistanc
725	32	39.0	246	2	D96491	hypothenical prote	798	32	39.0	397	2	G85935	probable resistanc
726	32	39.0	251	2	T04866	hypothenical prote	799	32	39.0	397	2	D65066	hypothenical prote
727	32	39.0	261	2	S78245	thiamin biosynthes	800	32	39.0	398	2	T46312	hypothenical prote
728	32	39.0	265	2	A72756	probable multipe	801	32	39.0	409	2	F90825	probable integrase
729	32	39.0	272	2	T33191	hypothenical prote	802	32	39.0	411	2	H83788	involved in spore
730	32	39.0	273	2	AB2811	conserved hypothen	803	32	39.0	412	2	G96816	hypothenical prote
731	32	39.0	273	2	E97589	hypothenical prote	804	32	39.0	414	2	B96808	protein F28K19.2 [
732	32	39.0	275	2	T37304	ras GTPase-activat	805	32	39.0	415	2	E83377	probable alcohol d
733	32	39.0	276	2	AE1141	hypothenical prote	806	32	39.0	416	2	A85684	probable integrase
734	32	39.0	276	2	C82084	conserved hypothen	807	32	39.0	418	1	A53888	thermolabile hemol
735	32	39.0	278	2	T29994	hypothenical prote	808	32	39.0	418	2	H82485	NADH dehydrogenase
736	32	39.0	279	2	F86842	prephenate dehydra	809	32	39.0	419	2	D86728	hypothenical prote
737	32	39.0	279	2	S52582	probable imidazole	810	32	39.0	421	2	T04753	hypothenical prote
738	32	39.0	281	2	T065530	agmatinase (EC 3.5	811	32	39.0	423	2	H83603	hypothenical prote
739	32	39.0	284	1	F64338	glycosyl transfera	812	32	39.0	424	2	H87558	cytochrome P450 fa
740	32	39.0	290	2	H64431	transcription regu	813	32	39.0	424	2	H89308	protein F14H8.1 [i
741	32	39.0	293	2	B98866	hypothenical prote	814	32	39.0	424	2	S17571	carboxypeptidase T
742	32	39.0	294	2	D90194	hypothenical prote	815	32	39.0	428	2	C75110	translation elonga
743	32	39.0	294	2	T45662	hypothenical prote	816	32	39.0	428	2	B75051	hypothenical prote
744	32	39.0	295	2	AD1890	hypothenical prote	817	32	39.0	428	2	G71023	translation elonga
745	32	39.0	300	2	F83804	cation efflux syst	818	32	39.0	433	2	T39745	hypothenical prote
746	32	39.0	301	2	JW0078	amine sulfotransfer	819	32	39.0	439	2	H83699	sodium-dependent c
747	32	39.0	302	2	AH0191	probable polysacch	820	32	39.0	443	2	T08136	probable omega-6 d
748	32	39.0	302	2	H83478	probable binding p	821	32	39.0	445	2	S73996	MG148 homolog Vxps
749	32	39.0	302	2	S27846	hypothenical prote	822	32	39.0	445	2	A80359	hypothenical prote
750	32	39.0	304	2	AG1837	WD-40 repeat prote	823	32	39.0	448	2	D85362	hypothenical prote
751	32	39.0	307	2	A71057	probable sulfatase	824	32	39.0	450	2	AF0987	glutathione-disulf
752	32	39.0	308	2	T21874	hypothenical prote	825	32	39.0	453	2	E96651	protein T3P18.13 [
753	32	39.0	309	1	B43311	sulfur oxygenase/r	826	32	39.0	454	2	JC4886	bleomycin hydrolas
754	32	39.0	309	2	T20518	hypothenical prote	827	32	39.0	456	2	S61971	hypothenical prote
755	32	39.0	310	2	B65094	hypothenical prote	828	32	39.0	458	2	F86433	protein T17H7.5 [i
756	32	39.0	310	2	F85966	probable transcrip	829	32	39.0	459	2	AI3384	biocin carboxylase
757	32	39.0	310	2	G91121	probable transcrip	830	32	39.0	459	2	S13064	1D-myo-inositol-tr
758	32	39.0	311	2	C72381	hypothenical prote	831	32	39.0	461	2	JN0129	1D-myo-inositol-tr
759	32	39.0	314	2	I50811	MHC class I protei	832	32	39.0	464	2	T01957	hypothenical prote

833	32	39.0	466	2	S62330	beta-fructofuranos	906	32	39.0	690	2	E86442	probable PPR-repea
834	32	39.0	470	2	T49272	hypothetical prote	907	32	39.0	692	2	A97013	hypothetical prote
835	32	39.0	471	2	T20938	hypothetical prote	908	32	39.0	693	2	T00256	hypothetical prote
836	32	39.0	477	2	A56449	protoporphyrinogen	909	32	39.0	694	2	I40866	exo-alpha-stalidas
837	32	39.0	481	2	D64883	Aminobenzoyl-gluta	910	32	39.0	695	2	H86900	DNA mismatch repai
838	32	39.0	481	2	A90869	hypothetical prote	911	32	39.0	698	2	T21781	hypothetical prote
839	32	39.0	481	2	H85749	hypothetical prote	912	32	39.0	701	2	S64599	probable membrane
840	32	39.0	486	2	T15281	hypothetical prote	913	32	39.0	702	2	S12638	transposition prot
841	32	39.0	488	2	T42038	catalase [EC 1.11.	914	32	39.0	721	2	T45495	probable transposa
842	32	39.0	491	2	JC7169	54K polar flagella	915	32	39.0	727	2	AH2134	DNA topoisomerase
843	32	39.0	491	2	E95356	probable ABC trans	916	32	39.0	731	1	S20687	DNA ligase (NAD) (
844	32	39.0	492	1	S68856	cytochrome P450 2L	917	32	39.0	746	2	A84800	hypothetical prote
845	32	39.0	492	2	A61382	phosphorylation re	918	32	39.0	751	2	T27691	hypothetical prote
846	32	39.0	493	2	E35115	anthranilate synth	919	32	39.0	752	2	S51866	HPRL protein - yea
847	32	39.0	493	2	T23030	hypothetical prote	920	32	39.0	755	2	G86469	protein Fl2k21.6 (
848	32	39.0	495	2	S43294	bone morphogenetic	921	32	39.0	755	2	A81436	probable outer mem
849	32	39.0	498	2	T51430	dolichyl-phosphate	922	32	39.0	761	2	T41304	probable rna-bindi
850	32	39.0	500	2	D87541	beta-xylosidase [i	923	32	39.0	763	2	I50807	complement factor
851	32	39.0	504	2	C85485	probable carnitine	924	32	39.0	767	1	CO2PCD	cdc10 start contro
852	32	39.0	504	2	C90634	probable carnitine	925	32	39.0	772	2	S62481	hypothetical prote
853	32	39.0	504	2	H64724	probable carnitine	926	32	39.0	775	2	T40647	ubiquitin carboxyl
854	32	39.0	505	2	I38396	protein-tyrosine k	927	32	39.0	777	2	T38769	hypothetical prote
855	32	39.0	505	2	A39128	anthranilate synth	928	32	39.0	786	2	D75630	glycerophosphate kina
856	32	39.0	505	2	B84831	hypothetical prote	929	32	39.0	789	2	B87461	polycerophosphoryl
857	32	39.0	512	2	H96759	probable steroid 2	930	32	39.0	790	2	T19040	hypothetical prote
858	32	39.0	512	2	I49552	protein-tyrosine k	931	32	39.0	794	2	T52441	hypothetical prote
859	32	39.0	512	2	H86206	hypothetical prote	932	32	39.0	797	2	D86247	hypothetical prote
860	32	39.0	515	2	S02194	DNA-directed RNA p	933	32	39.0	808	2	T51138	probable glutamate
861	32	39.0	516	2	G70149	hypothetical prote	934	32	39.0	823	2	B70203	ATP-dependent heli
862	32	39.0	517	2	T02403	probable beta-gluc	935	32	39.0	824	2	T20351	hypothetical prote
863	32	39.0	518	2	S42387	MIPP protein homol	936	32	39.0	827	2	F72414	ribonucleotide red
864	32	39.0	522	2	D72349	conserved hypotet	937	32	39.0	829	2	G86763	DNA topoisomerase
865	32	39.0	522	2	S63057	hypothetical prote	938	32	39.0	867	2	E86815	ClpB protein [mpo
866	32	39.0	531	2	F69949	conserved hypotet	939	32	39.0	874	2	B86322	FlpA1.8 protein -
867	32	39.0	542	1	QREBOA	oligopeptide-bindi	940	32	39.0	878	2	A83748	endo-beta-N-acetyl
868	32	39.0	542	2	B82910	CTP synthetase U02	941	32	39.0	899	2	B87553	DNA topoisomerase
869	32	39.0	543	1	F64871	oligopeptide-bindi	942	32	39.0	908	2	JN0819	transferrin-bindin
870	32	39.0	543	2	F85704	hypothetical prote	943	32	39.0	911	2	T08108	nitrate reductase
871	32	39.0	543	2	G30846	hypothetical prote	944	32	39.0	913	2	T18503	hypothetical prote
872	32	39.0	545	1	S44486	indole-3-pyruvate	945	32	39.0	917	2	E96807	nitrate reductase
873	32	39.0	556	2	T38479	myb-like DNA-bindi	946	32	39.0	917	2	S35228	nitrate reductase
874	32	39.0	557	2	JC5487	cellulase [EC 3.2.	947	32	39.0	918	2	T38786	translation initiat
875	32	39.0	562	2	AF0852	secretory protein	948	32	39.0	920	2	T18852	hypothetical prote
876	32	39.0	563	1	C2CLEM	cellulase [EC 3.2.	949	32	39.0	924	2	S34926	hypothetical prote
877	32	39.0	563	2	B70682	probable nitrite r	950	32	39.0	942	2	T39624	6-phosphofructokin
878	32	39.0	573	2	B70047	two-component sens	951	32	39.0	947	2	H85088	hypothetical prote
879	32	39.0	579	2	H88478	protein F47D12.7 [	952	32	39.0	949	2	S54020	probable membrane
880	32	39.0	582	2	AG0650	periplasmic oligop	953	32	39.0	955	2	T21612	hypothetical prote
881	32	39.0	583	2	S50959	probable membrane	954	32	39.0	956	2	T19046	ras GTPase-activat
882	32	39.0	583	2	H69165	hypothetical prote	955	32	39.0	963	2	S45167	chitin synthase (E
883	32	39.0	585	1	A41292	glutamate decarbox	956	32	39.0	971	2	A35697	transcription fact
884	32	39.0	585	1	JH0423	glutamate decarbox	957	32	39.0	987	2	H81722	polymorphic membra
885	32	39.0	585	1	S38533	glutamate decarbox	958	32	39.0	996	2	D86872	beta-galactosidase
886	32	39.0	585	2	JC4064	glutamate decarbox	959	32	39.0	1000	2	F70368	beta-galactosidase
887	32	39.0	586	2	AH2133	ATP-binding protei	960	32	39.0	1005	2	T31333	beta-galactosidase
888	32	39.0	587	2	C86744	myosin-crossreacti	961	32	39.0	1007	2	A30093	beta-D-galactosida
889	32	39.0	591	2	S51303	hypothetical prote	962	32	39.0	1013	2	G71460	probable outer mem
890	32	39.0	599	2	H72336	conserved hypotet	963	32	39.0	1015	2	I39697	beta-galactosidase
891	32	39.0	604	2	T37870	RNA-binding / Ran	964	32	39.0	1019	2	C96519	probable disease r
892	32	39.0	606	2	D86443	probable PPR-repea	965	32	39.0	1024	1	GBEC	beta-galactosidase
893	32	39.0	611	2	A83926	hypothetical prote	966	32	39.0	1024	2	E90678	beta-D-galactosida
894	32	39.0	626	2	A82771	hypothetical prote	967	32	39.0	1024	2	A85529	beta-D-galactosida
895	32	39.0	626	2	A42891	beta-galactosidase	968	32	39.0	1025	1	JC1266	beta-galactosidase
896	32	39.0	628	2	A51780	NADH dehydrogenase	969	32	39.0	1034	2	T30551	beta-galactosidase
897	32	39.0	628	2	AF1404	NADH dehydrogenase	970	32	39.0	1034	2	T30574	beta-galactosidase
898	32	39.0	642	2	E70683	probable helix-tur	971	32	39.0	1034	2	A24925	beta-galactosidase
899	32	39.0	645	2	C94999	ATP-dependent DNA	972	32	39.0	1042	1	GBECE	beta-galactosidase
900	32	39.0	660	2	H82281	conserved hypotet	973	32	39.0	1042	2	E85968	evolved beta-D-gal
901	32	39.0	662	2	A37226	glucose transport	974	32	39.0	1042	2	F91123	ras GTPase-activat
902	32	39.0	664	2	A33545	Na+/glucose cotran	975	32	39.0	1048	2	T15045	WD repeat domain p
903	32	39.0	664	2	C71106	hypothetical prote	976	32	39.0	1051	2	T48933	beta-galactosidase
904	32	39.0	671	2	S61099	leukotriene-A4 hyd	977	32	39.0	1060	2	AI0201	beta-galactosidase
905	32	39.0	675	1	FVFFB	brown protein - fr	978	32	39.0	1067	2	D96545	probable DNA polym

979 32 39.0 1080 2 A35088 phycobilisome link  
 980 32 39.0 1085 2 S55352 IFH1 protein - yea  
 981 32 39.0 1087 2 F72283 beta-galactosidase  
 982 32 39.0 1093 2 T51503 valine-tRNA ligase  
 983 32 39.0 1102 2 S44772 C29E4.4 protein -  
 984 32 39.0 1127 2 S47445 MDML protein - yea  
 985 32 39.0 1137 2 D89610 ras GTPase-activat  
 986 32 39.0 1140 1 I38908 UV-damaged DNA-bin  
 987 32 39.0 1140 1 S38777 UV-damaged DNA-bin  
 988 32 39.0 1140 2 JC7152 probable ATP-depen  
 989 32 39.0 1152 2 H88533 probable chemotaxi  
 990 32 39.0 1161 2 H95903 ras GTPase-activat  
 991 32 39.0 1170 2 T19042 hypothetical prote  
 992 32 39.0 1182 2 T46428 phosphatidylinosit  
 993 32 39.0 1188 2 JC4889 SH2-containing ino  
 994 32 39.0 1189 2 JC6118 ras GTPase-activat  
 995 32 39.0 1207 2 T19041 adenosine deaminas  
 996 32 39.0 1226 1 S65593 PiliY protein homo  
 997 32 39.0 1230 2 F82857 hypothetical prote  
 998 32 39.0 1231 2 T24415 adenylylate cyclase  
 999 32 39.0 1248 2 A53598 type VIII adenylyl  
 1000 32 39.0 1251 2 S48687

ALIGNMENTS

RESULT 1  
 A32576  
 beta-glucuronidase (BC 3.2.1.31) allele B precursor - mouse  
 N;Alternate names: beta-D-glucuronoside glucuronohydrolase  
 C;Species: Mus musculus (house mouse)  
 C;Date: 12-Oct-1989 #sequence\_revision 12-Oct-1989 #text\_change 09-Jul-2004  
 C;Accession: A32576; J;Gallagher, P.M.; D'Amore, M.A.; Carter, J.E.; Lund, S.D.; Rinchik, E.  
 R;Wawrzyniak, C.J.; Gallagher, P.M.; D'Amore, M.A.; Carter, J.E.; Lund, S.D.; Rinchik, E.  
 Mol. Cell. Biol. 9, 4074-4078, 1989  
 A;Title: DNA determinants of structural and regulatory variation within the murine beta-  
 A;Reference number: A32576; MUID:89384641; PMID:2779578  
 A;Accession: A32576  
 A;Molecule type: mRNA  
 A;Residues: 1-648 <WAW>  
 A;Cross-references: UNIPARC:UPI000016CDB0; GB:M28541; NID:g193720; PIDN:AAA63308.1; PID:  
 A;Experimental source: allele B  
 R;Funkenstein, B.; Leary, S.L.; Stein, J.C.; Catterall, J.F.  
 Mol. Cell. Biol. 8, 1160-1168, 1988  
 A;Title: Genomic organization and sequence of the Gus-s-a allele of the murine beta-gluc  
 A;Reference number: 149692; MUID:88216590; PMID:2835664  
 A;Accession: 149692  
 A;Status: translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-232, 'T', 234-264, 'D', 266-319, 'V', 321-427, 'K', 429-615, 'L', 617-648 <RES>  
 A;Cross-references: UNIPARC:UPI000016CDB5; GB:M19279; NID:g193524; PIDN:AAA37697.1; PID:  
 A;Experimental source: allele A  
 R;D'Amore, M.A.; Gallagher, P.M.; Korfhagen, T.R.; Ganschow, R.E.  
 Biochemistry 27, 7131-7140, 1988  
 A;Title: Complete sequence and organization of the murine beta-glucuronidase gene.  
 A;Reference number: A28954; MUID:89062453; PMID:3196706  
 A;Accession: A28954  
 A;Molecule type: DNA  
 A;Residues: 1-264, 'D', 266-319, 'V', 321-648 <DAM>  
 A;Cross-references: UNIPARC:UPI000027ABB; GB:J02836; GB:J03035; GB:M20204; NID:g193716;  
 R;Gallagher, P.M.; D'Amore, M.A.; Lund, S.D.; Ganschow, R.E.  
 Genomics 2, 215-219, 1988  
 A;Title: The complete nucleotide sequence of murine beta-glucuronidase mRNA and its dedu  
 A;Reference number: A29977; MUID:88284700; PMID:3397060  
 A;Accession: A29977  
 A;Molecule type: mRNA  
 A;Residues: 1-264, 'D', 266-319, 'V', 321-648 <GAL>

A;Cross-references: UNIPARC:UPI0000027ABB; GB:J03047; NID:g193522; PIDN:AAA37696.1; PID:  
 J;Li, H.; Takeuchi, K.H.; Manly, K.; Chapman, V.; Swank, R.T.  
 J. Biol. Chem. 265, 14732-14735, 1990  
 A;Title: The propeptide of beta-glucuronidase. Further evidence of its involvement in con  
 the serpin superfamily.  
 A;Reference number: A35798; MUID:90368633; PMID:2394691  
 A;Accession: A35798  
 A;Status: not compared with conceptual translation  
 A;Molecule type: mRNA  
 A;Residues: 593-648 <LIA>  
 A;Cross-references: UNIPARC:UPI0000175B40  
 A;Note: the location of the propeptide cleavage site was not demonstrated directly but re  
 C;Comment: in some tissues, a portion of this enzyme is retained in the endoplasmic retic  
 syn.  
 C;Genetics:  
 A;Gene: Gus  
 A;Map position: 5  
 A;Introns: 70/3; 132/3; 193/2; 241/1; 303/3; 351/3; 411/2; 460/2; 488/3; 547/3; 593/1  
 C;Superfamily: beta-glucuronidase  
 C;Keywords: glycosidase; hydrolase; lysosome  
 F;1-22/Domain: signal sequence #status predicted <SIG>  
 F;23-648/Product: beta-glucuronidase, ER-retained form #status predicted <ERMT>  
 F;23-633/Product: beta-glucuronidase, lysosomal (default) form #status predicted <LMAT>  
 F;634-648/Domain: carboxyl-terminal propeptide #status predicted <CPRO>

Query Match 54.9%; Score 45; DB 2; Length 648;  
 Best Local Similarity 57.1%; Pred. No. 7.5;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKKA 15  
 | : |||: |||  
 Db 442 WSVANEPSSALKPA 455

RESULT 2  
 T49366  
 myocyte-specific enhancer factor 2d related protein [imported] - Neurospora crassa  
 N;Alternate names: protein B1D1.200  
 C;Species: Neurospora crassa  
 C;Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 09-Jul-2004  
 C;Accession: T49366  
 R;Schulte, U.; Aign, V.; Hohseil, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,  
 submitted to the Protein Sequence Database, May 2000  
 A;Reference number: Z25022  
 A;Accession: T49366  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-624 <SCH>  
 A;Cross-references: UNIPROT:Q9PGA0; UNIPARC:UPI000006A730; EMBL:AL355927; GSPDB:GN00116;  
 A;Experimental source: BAC clone B1D1; strain OR74A  
 C;Genetics:  
 A;Gene: NCSP:B1D1.200  
 A;Map position: 6  
 A;Introns: 11/1; 148/1; 167/2; 244/1

Query Match 52.4%; Score 43; DB 2; Length 624;  
 Best Local Similarity 66.7%; Pred. No. 17;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNEPN 9  
 : |||: |||  
 Db 238 TWQLGNSPN 246

RESULT 3  
 CCR7P  
 cytochrome c2 - Rhodopseudomonas palustris (strain 2.1.37)  
 C;Species: Rhodopseudomonas palustris  
 C;Date: 22-May-1981 #sequence\_revision 22-May-1981 #text\_change 31-Dec-2004  
 C;Accession: A00083  
 R;Ambler, R.P.; Daniel, M.; Hermoso, J.; Meyer, T.E.; Bartsch, R.G.; Kamen, M.D.  
 Nature 278, 659-660, 1979  
 A;Title: Cytochrome c-2 sequence variation among the recognised species of purple nonsulf

A;Reference number: A00086; MUID:79199667; PMID:221822

A;Accession: A00083

A;Molecule type: protein

A;Residues: 1-114 <AMB>

A;Cross-references: UNIPROT:P00091; UNIPARC:UPI0000128855

A;Experimental source: ATCC 17007

C;Superfamily: cytochrome c/cytochrome c2; cytochrome c homology

C;Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; photosynthesis

F;4-110/Domain: cytochrome c homology <CYC>

F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

F;13,16/Binding site: heme (Cys) (covalent) #status predicted

F;17,93/Binding site: heme iron (His, Met) (axial ligands) #status predicted

Query Match 51.2%; Score 42; DB 1; Length 114;

Best Local Similarity 77.8%; Pred. No. 3.6;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 6 NEPNSFLKK 14

Db 68 NDPNAFLKK 76

#### RESULT 4

S48273

probable transcription factor YBR108w - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: hypothetical protein YBR0901

C;Species: *Saccharomyces cerevisiae*

C;Date: 01-Aug-1995 #sequence\_revision 01-Sep-1995 #text\_change 09-Jul-2004

C;Accession: S48273; S45976; S44688

R;Mannhaupt, G.; Stucka, R.; Ehnlé, S.; Vetter, I.; Feldmann, H.

Yeast 10, 1363-1381, 1994

A;Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.

A;Reference number: S48255; MUID:95208357; PMID:7900426

A;Accession: S48273

A;Status: nucleic acid sequence not shown

A;Molecule type: DNA

A;Residues: 1-848 <MAN>

A;Cross-references: UNIPROT:P38266; UNIPARC:UPI000013A3E4; EMBL:X78993; NID:g476045; PID

R;Feldmann, H.; Mannhaupt, G.; Schwarzlöse, C.; Vetter, I.

submitted to the Protein Sequence Database, August 1994

A;Reference number: S45927

A;Accession: S45976

A;Molecule type: DNA

A;Residues: 1-848 <FE2>

A;Cross-references: UNIPARC:UPI000013A3E4; EMBL:Z35977; NID:g536378; PID:g536379; MIPS:Y

C;Genetics:

A;Cross-references: SGD:S0000312

A;Map position: 2R

Query Match

Best Local Similarity 51.2%; Score 42; DB 2; Length 848;

Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 4 LGNEPNSFLKK 14

Db 715 LGDAPNSFIRK 725

#### RESULT 5

T08469

endo-1,4-beta-xylanase (EC 3.2.1.8) - *Dictyoglomus thermophilum*

C;Species: *Dictyoglomus thermophilum*

C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 31-Dec-2004

C;Accession: T08469

R;Gibbs, M.D.; Reeves, R.A.; Bergquist, P.L.

Appl. Environ. Microbiol. 61, 4403-4408, 1995

A;Title: Cloning, sequencing, and expression of a xylanase gene from the extreme thermop

A;Reference number: Z16432; MUID:96086022; PMID:8534104

A;Accession: T08469

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-352 <GIB>

A;Cross-references: UNIPROT:Q12603; UNIPARC:UPI000013907F; EMBL:L39866; NID:g973982; PID

A;Experimental source: strain Rt46B.1

C;Genetics:

A;Note: xyna

C;Function:

A;Description: hydrolyzes xylan to xylotriose and xylobiose but could not hydrolyze xylo

C;Superfamily: xylanase; Streptomyces endo-1,4-beta-xylanase A homology

C;Keywords: glycosidase; hydrolase; polysaccharide degradation

F;60-352/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 50.6%; Score 41.5; DB 2; Length 352;

Best Local Similarity 40.0%; Pred. No. 16;

Matches 8; Conservative 5; Mismatches 2; Indels 5; Gaps 1;

Qy 1 SWELGNE-----PNSFLKKA 15

Db 149 AWDVVNEALSNDPNEFLRRA 168

#### RESULT 6

B83007

conserved hypothetical protein PA5115 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C;Species: *Pseudomonas aeruginosa*

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004

C;Accession: B83007

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bri

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathog

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: B83007

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-194 <STO>

A;Cross-references: UNIPROT:Q9HU69; UNIPARC:UPI000000CSF0C; GB:AE004924; GB:AE004091; NID

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA5115

Query Match

Best Local Similarity 50.0%; Score 41; DB 2; Length 194;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 5 GNEPNSFLKK 14

Db 14 GTEPNAFLKE 23

#### RESULT 7

T19330

hypothetical protein C16C10.9 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T19330

R;Lloyd, C.

submitted to the EMBL Data Library, November 1994

A;Reference number: Z19108

A;Accession: T19330

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-274 <WIL>

A;Cross-references: UNIPROT:Q09465; UNIPARC:UPI0000060EFC; EMBL:Z46787; PIDN:CAA86747.1;

A;Experimental source: clone C16C10

C;Genetics:

A;Gene: CESP:C16C10.9

A;Map position: 3

A;Introns: 88/3; 117/3; 143/3; 170/2; 193/3; 222/3; 248/3

C;Superfamily: *Caenorhabditis elegans* hypothetical protein C16C10.9

Query Match

Best Local Similarity 50.0%; Score 41; DB 2; Length 274;

Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKK 14  
:|||:|  
Db 173 FEMNGDPFFVK 185

## RESULT 8

B84900  
hypothetical protein At2g46220 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: B84900  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: B84900  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-550 <STO>  
A:Cross-references: UNIPROT:O82347; UNIPARC:UPI000017A71B; GB:AE002093; NID:g3702327; PI  
C:Genetics:  
A:Gene: At2g46220  
A:Map position: 2

Query Match 50.0%; Score 41; DB 2; Length 550;

Best Local Similarity 61.5%; Pred. No. 33;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLK 13  
|||:|  
Db 434 SWHLGSETKTLK 446

## RESULT 9

A25047  
beta-glucuronidase (BC 3.2.1.31) precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 30-Jun-1998 #sequence\_revision 30-Jun-1998 #text\_change 09-Jul-2004  
C:Accession: A25047; S00345  
R;Nishimura, Y.; Rosenfeld, M.G.; Kreibich, G.; Gubler, U.; Sabatini, D.D.; Adesnik, M.;  
Proc. Natl. Acad. Sci. U.S.A. 83, 7292-7296, 1986  
A:Title: Nucleotide sequence of rat preputial gland beta-glucuronidase cDNA and in vitro  
A:Reference number: A25047; MUID:87016933; PMID:3463967  
A:Accession: A25047  
A:Molecule type: mRNA  
A:Residues: 1-648 <NIS>  
A:Cross-references: UNIPROT:P06760; UNIPARC:UPI0000126911; GB:M13962; NID:g204329; PIDN:  
A:Experimental source: female preputial gland  
R;Powell, P.P.; Kyle, J.W.; Miller, R.D.; Pantano, J.; Grubb, J.H.; Sly, W.S.  
Biochem. J. 250, 547-555, 1988  
A:Title: Rat liver beta-glucuronidase. cDNA cloning, sequence comparisons and expression  
A:Reference number: S00345; MUID:88183378; PMID:3355537  
A:Accession: S00345  
A:Molecule type: mRNA  
A:Residues: 'E', 15-20, 'L', 22-486, 'L', 488-648 <POW>  
A:Cross-references: UNIPARC:UPI0000017097A; EMBL:F00717; NID:g56270; PIDN:CAA68705.1; PID  
C:Superfamily: beta-glucuronidase  
C:Keywords: glycosidase; hydrolase  
F;1-22/Domain: signal sequence #status predicted <SIG>  
F;23-648/Product: beta-glucuronidase #status predicted <MAT>

Query Match 50.0%; Score 41; DB 2; Length 648;

Best Local Similarity 57.1%; Pred. No. 40;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKKA 15  
|:|:|  
Db 442 WSVANEPVSSLKPA 455

## RESULT 10

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AC2609  
A>Status: preliminary

## C36955

flagellar biosynthesis-specific protein fliQ - Yersinia pestis plasmid pCD1  
C:Species: Yersinia pestis  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: C36955  
R;Fields, K.A.; Plano, G.V.; Straley, S.C.  
J. Bacteriol. 176, 569-579, 1994

A:Title: A low-Ca(2+) response (LCR) secretion (ysc) locus lies within the lcrB region of

A:Reference number: A36955; MUID:94131934; PMID:8300512

A:Accession: C36955

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-79 <PIE>

A:Cross-references: UNIPROT:P40298; UNIPARC:UPI000016E1D4; GB:L22495; NID:g2465655; PIDN:

C:Genetics:

A:Gene: yscS

A:Genome: plasmid

C:Superfamily: flagellar biosynthesis-specific protein

Query Match 49.4%; Score 40.5; DB 1; Length 79;

Best Local Similarity 64.3%; Pred. No. 4.4;

Matches 9; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

QY 1 SWELGNEPNSFLKK 14  
|||:|  
Db 67 SW-LGNELHSFVQK 79

## RESULT 11

## B97391

hypothetical protein AGR\_C\_459 [imported] - Agrobacterium tumefaciens (strain C58, Cereor

C:Species: Agrobacterium tumefaciens

C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Jul-2004

C:Accession: B97391

R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ourollo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; MUID:21608551; PMID:11743194

A:Accession: B97391

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-117 <KUR>

A:Cross-references: UNIPROT:Q8UIM6; UNIPARC:UPI00000D175E; GB:AE007869; PIDN:AAK86083.1;

C:Genetics:

A:Gene: AGR\_C\_459

A:Map position: circular chromosome

Query Match 48.8%; Score 40; DB 2; Length 117;

Best Local Similarity 52.6%; Pred. No. 8.4;

Matches 10; Conservative 1; Mismatches 4; Indels 4; Gaps 1;

QY 1 SWEL-----GNPNSFLKKA 15  
||:| | | | | | | |  
Db 47 SWQLNRNGAELNGFLKNA 65

## RESULT 12

## AC2609

hypothetical protein Atu0267 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 09-Jul-2004

C:Accession: AC2609

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan,

A.; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, J.

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AC2609

A>Status: preliminary

```

A:Molecule type: DNA
A:Residues: 1-117 <KUR>
A:CROSS-references: UNIPROT:Q8UIM6; UNIPARC:UPI00000D175E; GB:AE008688; PIDN:AAL41289.1;
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu0267
A:Map position: circular chromosome

Query Match      48.8%; Score 40; DB 2; Length 117;
Best Local Similarity 52.6%; Pred. No. 8.4;
Matches 10; Conservative 1; Mismatches 4; Indels 4; Gaps 1;

QY 1 SWELGNEPNSFLKKA 15
    |||:|:|:|:|:|:|
Db 47 SWQLNRSNGAEINFLKVA 65

RESULT 13
T32412
hypothetical protein W09B6.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T32412
R:Goela, D.; Maggi, L.; Andrews, S.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid W09B6.
A:Reference number: Z21162
A:Accession: T32412
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-270 <GOE>
A:CROSS-references: UNIPROT:Q9GZI4; UNIPARC:UPI0000080DB9; EMBL:AF025469; PIDN:AAB71052.
A:Experimental source: strain Bristol N2; clone W09B6
C:Genetics:
A:Gene: CESP:W09B6.5
A:Map position: 2
A:Introns: 43/3; 85/3; 141/2; 192/1; 264/1
C:Superfamily: Caenorhabditis elegans hypothetical protein W09B6.5

Query Match      48.8%; Score 40; DB 2; Length 270;
Best Local Similarity 47.1%; Pred. No. 22;
Matches 8; Conservative 3; Mismatches 4; Indels 2; Gaps 1;

QY 1 SWELGNE--PNSFLKKA 15
    :|||:|:|:|:|:|
Db 59 NWTIGTDFVPNFFRKA 75

RESULT 14
S54263
rep A protein - Bacteroides fragilis
C:Species: Bacteroides fragilis
C:Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Jul-2004
C:Accession: S54263
R:Haggoud, A.; Trinh, S.; Mohieddine, M.; Reysset, G.
submitted to the EMBL Data Library, April 1995
A:Description: Genetic analysis of the minimal replicon of plasmid pIP417 and comparison
A:Reference number: S54263
A:Accession: S54263
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-331 <HAG>
A:CROSS-references: UNIPROT:Q45152; UNIPARC:UPI00000B0F0D; EMBL:X86702; NID:g804960; PID

Query Match      48.8%; Score 40; DB 2; Length 331;
Best Local Similarity 46.7%; Pred. No. 28;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
    |||:|:|:|:|:|:|
Db 252 SWLSLSVRSYLRKNS 266
```

```

RESULT 15
B27539
variant surface glycoprotein YNat 1.3 precursor - Trypanosoma congolense
C:Species: Trypanosoma congolense
C:Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jul-2004
C:Accession: B27539
R:Strickler, J.E.; Binder, D.A.; L'Italien, J.J.; Shimamoto, G.T.; Wait, S.W.; Dalheim, I.
Biochemistry 26, 796-805, 1987
A:Title: Trypanosoma congolense: structure and molecular organization of the surface glycoprotein
A:Reference number: A90525; MUID:87185370; PMID:3567147
A:Molecule type: mRNA
A:Residues: 1-413 <STR>
A:CROSS-references: UNIPROT:P20949; UNIPARC:UPI0000138E48; GB:M15113; NID:g162439; PID:g
C:Keywords: glycoprotein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-413/Product: variant surface glycoprotein YNat 1.3 #status predicted <MAT>

Query Match      48.8%; Score 40; DB 2; Length 413;
Best Local Similarity 53.8%; Pred. No. 36;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKK 14
    |||:|:|:|:|:|
Db 60 WALGENPTSRIKK 72

RESULT 16
A84709
cinnamate-4-hydroxylase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: A84709
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84709
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-505 <STO>
A:CROSS-references: UNIPROT:P92994; UNIPARC:UPI0000000F6E; GB:AE002093; NID:g1946370; PI
C:Genetics:
A:Gene: At2g30490
A:Map position: 2
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C:Keywords: heme; iron; metalloprotein
F:447/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match      48.8%; Score 40; DB 2; Length 505;
Best Local Similarity 53.8%; Pred. No. 45;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLK 13
    :|||:|:|:|:|:|
Db 400 AWWLANNPNSWKK 412

RESULT 17
F81191
hypothetical protein NMB0511 [imported] - Neisseria meningitidis (strain MC58 serogroup F
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C:Accession: F81191
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
```

```
A;Accession: F81191
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-640 <TET>
A;Cross-references: UNIPROT:Q9K0R9; UNIPARC:UPI0000032FA8; GB:AE002098; NID:
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB0511

Query Match      48.8%; Score 40; DB 2; Length 640;
Best Local Similarity 50.0%; Pred. No. 59;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKKA 15
   ||| ||| |||
Db 535 WVLNVSPNDYLKEA 548

RESULT 18
A26581
beta-glucuronidase (EC 3.2.1.31) precursor - human
C;Species: Homo sapiens (man)
C;Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 09-Jul-2004
C;Accession: A26581; A40337; A24983; A36538
R;Oshima, A.; Kyle, J.W.; Miller, R.D.; Hoffmann, J.W.; Powell, P.P.; Grubb, J.H.; Sly,
Proc. Natl. Acad. Sci. U.S.A. 84, 685-689, 1987
A;Title: Cloning, sequencing, and expression of cDNA for human beta-glucuronidase.
A;Reference number: A26581; MUID:87118233; PMID:3468507
A;Accession: A26581
A;Molecule type: mRNA
A;Residues: 1-651 <OSH>
A;Cross-references: UNIPROT:P08236; UNIPARC:UPI000003BC43; GB:M15182; NID:G183232; PIDN:
A;Experimental source: placenta
R;Shipley, J.M.; Miller, R.D.; Wu, B.M.; Grubb, J.H.; Christensen, S.G.; Kyle, J.W.; Sly
Genomics 10, 1009-1018, 1991
A;Title: Analysis of the 5' flanking region of the human beta-glucuronidase gene.
A;Reference number: A40337; MUID:92009900; PMID:1916806
A;Accession: A40337
A;Molecule type: DNA
A;Residues: 1-70 <SHI>
A;Cross-references: UNIPARC:UPI000016AA04; GB:M65002; NID:G183706; PIDN:AAA52622.1; PID:
R;Guise, K.S.; Korneluk, R.G.; Waye, J.; Lamhonwah, A.M.; Quan, F.; Palmer, R.; Ganschow
Gene 34, 105-110, 1985
A;Reference number: A24983; MUID:852332043; PMID:3924735
A;Accession: A24983
A;Molecule type: mRNA
A;Residues: 520-585 <GUI>
A;Cross-references: UNIPARC:UPI000016AA03; GB:M10618; NID:G183704; PIDN:AAA52621.1; PID:
R;Tomatsu, S.; Fukuda, S.; Sukegawa, K.; Ikeda, Y.; Yamada, S.; Yamada, Y.; Sasaki, T.;
Am J. Hum. Genet. 48, 89-96, 1991
A;Title: Mucopolysaccharidosis type VII: characterization of mutations and molecular het
A;Reference number: A36538; MUID:91090114; PMID:1702266
A;Accession: A36538
A;Molecule type: mRNA
A;Residues: 378-385 616-621 643-651 <TOM>
A;Cross-references: UNIPARC:UPI00000175B41
C;Genetics:
A;Gene: GDB:GUSB
A;Cross-references: GDB:120025; OMIM:253220
A;Map position: 7q22-7q22
C;Superfamily: beta-glucuronidase
C;Keywords: glycoprotein; glycosidase; homotetramer; hydrolase; lysosome
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-651/Product: beta-glucuronidase, placental #status predicted <MAT>

Query Match      48.8%; Score 40; DB 2; Length 651;
Best Local Similarity 50.0%; Pred. No. 60;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKKA 15
   ||| ||| |||
Db 446 WSVANEPASHLESA 459
```

```
RESULT 19
A35269
translation initiation factor IF-2 - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 17-Aug-1990 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C;Accession: A35269; B35269; S31994; G69644
R;Shazand, K.; Tucker, J.; Chiang, R.; Stansmore, K.; Sperling-Petersen, H.U.; Grunberg-
J. Bacteriol. 172, 2675-2687, 1990
A;Title: Isolation and molecular genetic characterization of the Bacillus subtilis gene
A;Reference number: A35269; MUID:90236932; PMID:2110148
A;Accession: A35269
A;Molecule type: DNA
A;Residues: 1-716 <SHA1>
A;Cross-references: UNIPROT:P17889; UNIPARC:UPI000012D2D0; GB:M34836; NID:G143358; PIDN:
A;Accession: B35269
A;Molecule type: DNA
A;Residues: 94-716 <SHA2>
A;Cross-references: UNIPARC:UPI0000174801
R;Shazand, K.
Submitted to the EMBL Data Library, November 1992
A;Reference number: S31990
A;Accession: S31994
A;Molecule type: DNA
A;Residues: 1-716 <SHA3>
A;Cross-references: UNIPARC:UPI000012D2D0; EMBL:Z18631; NID:G49314; PIDN:CAA79234.1; PID:
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
A.; Harwood, C.R.; Henaout, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;
Koeter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror,
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: G69644
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-55, 'A', 57-716 <KUN>
A;Cross-references: UNIPARC:UPI0000060405; GB:Z99112; GB:AL009126; NID:G26333902; PIDN:CAF
A;Experimental source: strain 168
C;Comment: IF-2, one of the essential components for the initiation of protein synthesis
mal subunits. It is also involved in the hydrolysis of GTP during the formation of the 70
C;Comment: Alternative initiation codons in the same reading frame do appear to be utiliz
C;Genetics:
A;Gene: infB
C;Superfamily: translation initiation factor IF-2; translation elongation factor Tu homol
C;Keywords: alternative initiators; GTP binding; nucleotide binding; P-loop; protein bios
F;1-716/Product: translation initiation factor IF-2-alpha #status predicted <ALP>
F;94-716/Product: translation initiation factor IF-2-beta #status predicted <BET>
F;220-329/Domain: translation elongation factor Tu homology <ETU>
F;226-233/Region: nucleotide-binding motif A (P-loop)
F;326-329/Region: GTP-binding NKXD motif
F;362-364/Region: GTP-binding SAK/L motif
F;232,233,235,326,327,329,362/Binding site: Mg-GTP (Lys, Thr, Thr, Asn, Lys, Asp, Ser) #
Query Match      48.8%; Score 40; DB 1; Length 716;
Best Local Similarity 58.3%; Pred. No. 67;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 ELGNEPNSFLKK 14
   ||| ||| |||
Db 151 ELGKEPSFLIKK 162

RESULT 20
```



D82704  
conserved hypothetical protein XF1252 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004  
C:Accession: D82704  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: D82704  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1641 <STM>  
A:Cross-references: UNIPROT:Q9PDX7; UNIPARC:UPI000013A4EB; GB:AE003959; GB:AE003849; NID  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
as-Neto, E.; Docena, C.; El-Dorzy, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A:Authors: Martins, E.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF1252

Query Match 48.8%; Score 40; DB 2; Length 1641;  
Best Local Similarity 70.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 SWELGNPNPS 10  
||||| |:  
Db 212 SWELSNQNGS 221

RESULT 21  
S48138  
ubiquinol-cytochrome-c reductase (EC 1.10.2.2) chain VIII - yeast (Saccharomyces cerevis  
N:Alternate names: protein J0526; protein YJL166w; ubiquinol-cytochrome-c reductase 11k  
C:Species: Saccharomyces cerevisiae  
C>Date: 30-Sep-1991 #sequence\_revision 11-Nov-1994 #text\_change 09-Jul-2004  
C:Accession: S48138; S05860; S48137; S56949  
R:Koerkamp, M.G.  
submitted to the EMBL Data Library, June 1993  
A:Reference number: S48138  
A:Accession: S48138  
A:Molecule type: DNA  
A:Residues: 1-94 <KOE>  
A:Cross-references: UNIPROT:P08525; UNIPARC:UPI000004F931; EMBL:X05550; NID:G312353; PID  
E:Maarse, A.C.; Grivell, L.A.  
Eur. J. Biochem. 165, 419-425, 1987  
A:Title: Nucleotide sequence of the gene encoding the 11-kDa subunit of the ubiquinol--c  
A:Reference number: S05860; MUID:87246620; PMID:3036507  
A:Accession: S05860  
A:Molecule type: DNA  
A:Residues: 1-61, 'LREFIGGRTVTSIMNFCTAKLVKSKWELMENIV' <MAA1>  
A:Cross-references: UNIPARC:UPI0000179F9B; EMBL:X05550  
A:Accession: S48137  
A:Molecule type: protein  
A:Residues: 2-11 <MAA2>  
A:Cross-references: UNIPARC:UPI0000179F9C  
R:Obermaier, B.; Piravandi, E.; Rinke, M.; Domdey, H.  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56937  
A:Accession: S56949  
A:Molecule type: DNA

A:Residues: 1-94 <OBE>  
A:Cross-references: UNIPARC:UPI000004F931; EMBL:Z49441; NID:G1008355; PID:G1008356; MIPS  
C:Genetics:  
A:Gene: SGD:QCR8  
A:Cross-references: SGD:S0003702; MIPS:YJL166w  
A:Map position: 10L  
A:Genome: nuclear  
C:Function:  
A:Description: oxidoreductase  
A:Pathway: oxidative phosphorylation; respiratory chain  
C:Superfamily: Schizosaccharomyces pombe ubiquinol-cytochrome-c reductase chain VIII  
C:Keywords: membrane-associated complex; mitochondrion; oxidative phosphorylation; oxid  
F:2-94/Product: ubiquinol-cytochrome-c reductase chain VIII #status experimental <MAT>

Query Match 47.6%; Score 39; DB 2; Length 94;  
Best Local Similarity 63.8%; Pred. No. 9.9;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 WELGNPNPSFL 12  
|: ||||| |:  
Db 70 WKNNGNEYNFL 80

RESULT 22  
T22536  
hypothetical protein F53B2.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T22536  
R:Smyle, R.  
submitted to the EMBL Data Library, June 1996  
A:Reference number: Z19577  
A:Accession: T22536  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-263 <WIL>  
A:Cross-references: UNIPROT:Q20705; UNIPARC:UPI0000061070; EMBL:Z73908; PIDN:CAA98129.1;  
A:Experimental source: clone F53B2  
C:Genetics:  
A:Gene: CESP:F53B2.4  
A:Map position: 4  
A:Introns: 15/3; 85/3; 138/3; 204/3  
C:Superfamily: Caenorhabditis elegans hypothetical protein C33A12.9b

Query Match 47.6%; Score 39; DB 2; Length 263;  
Best Local Similarity 53.8%; Pred. No. 32;  
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SWELGNPNPSFLK 13  
||| :||| |:  
Db 37 SWCASQQPNHFLK 49

RESULT 23  
AB4326  
hypothetical protein Vngl1740c [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: AB4326  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S  
; Leitthausner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Lif  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: AB4160; MUID:20504483; PMID:11016950  
A:Accession: AB4326  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-325 <STO>  
A:Cross-references: UNIPROT:Q9HP97; UNIPARC:UPI000006396C; GB:AE004437; NID:G10581200; P  
C:Genetics:  
A:Gene: VNG1740C

Query Match 47.6%; Score 39; DB 2; Length 325;  
 Best Local Similarity 58.3%; Pred. No. 41;  
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 ELGNEPNSFLKK 14  
 ||||| :|||:  
 Db 28 ELGHEDAFLKR 39

RESULT 24  
 T04763  
 chitinase homolog T16H5.180 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
 C:Accession: T04763  
 R:Bevan, M.; De Haan, M.; Maarse, A.C.; Grivell, L.A.; Bancroft, I.; Mewes, H.W.; Mayer, submitted to the Protein Sequence Database, June 1998  
 A:Reference number: Z15383  
 A:Accession: T04763  
 A:Molecule type: DNA  
 A:Residues: 1-366 <BV>  
 A:Cross-references: UNIPROT:O81863; UNIPARC:UPI00000A9E38; EMBL:AL024486  
 A:Experimental source: cultivar Columbia; BAC clone T16H5  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 230/1  
 A:Note: T16H5.180  
 C:Superfamily: Streptomyces chitinase chi40

Query Match 47.6%; Score 39; DB 2; Length 366;  
 Best Local Similarity 46.7%; Pred. No. 47;  
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15  
 ||| :|||:  
 Db 345 SWHIGADDNSRLSRA 359

RESULT 25  
 F87348  
 mannanase, probable [imported] - Caulobacter crescentus  
 C:Species: Caulobacter crescentus  
 C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
 C:Accession: F87348  
 R:Kierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J. B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A:Title: Complete Genome Sequence of Caulobacter crescentus.  
 A:Reference number: A87249; MUID:21173698; PMID:11259647  
 A:Accession: F87348  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-442 <STO>  
 A:Cross-references: UNIPROT:Q9AA10; UNIPARC:UPI00000C718D; GB:AE005673; NID:gi3422046; H C:Genetics:  
 A:Gene: CC0801

Query Match 47.6%; Score 39; DB 2; Length 442;  
 Best Local Similarity 75.0%; Pred. No. 59;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNEP 8  
 ||| :|||:  
 Db 216 SWQLANEP 223

RESULT 26  
 F70189  
 rod shape-determining protein (mreB-2) homolog - Lyme disease spirochete  
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
 C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 09-Jul-2004

C:Accession: F70189  
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B. Nature 390, 580-586, 1997  
 A:Authors: Smith, H.O.; Venter, J.C.  
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
 A:Reference number: A70100; MUID:98065943; PMID:9403685  
 A:Accession: F70189  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-459 <KLB>  
 A:Cross-references: UNIPROT:O51661; UNIPARC:UPI00000575B4; GB:AE001171; GB:AE000783; NID: A:Experimental source: strain B31  
 C:Superfamily: rod shape-determining protein

Query Match 47.6%; Score 39; DB 2; Length 459;  
 Best Local Similarity 50.0%; Pred. No. 61;  
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLK 13  
 ||| :|||:  
 Db 123 WKLGQPSSEFGK 134

RESULT 27  
 T40391  
 phosphoprotein phosphatase (EC 3.1.3.16) SPBC3F6.01c [similarity] - fission yeast (Schiz N:Alternate names: serine/threonine phosphatase PF5 homolog  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 28-Apr-2003  
 C:Accession: T40391  
 R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Churcher, C.M.  
 submitted to the EMBL Data Library, February 1998  
 A:Reference number: Z21925  
 A:Accession: T40391  
 A:Status: translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-473 <LYN>  
 A:Cross-references: UNIPARC:UPI000006B641; EMBL:AL022019; PIDN:CAAL7690.2; GSPDB:GN00067; A:Experimental source: strain 972h-; cosmid c3f6  
 C:Genetics:  
 A:Gene: SPDB:SPBC3F6.01c  
 A:Map position: 2  
 C:Superfamily: protein phosphatase 5; phosphoesterase core homology; phosphoprotein phosph C:Keywords: iron; metalloprotein; nucleus; phosphoric monoester hydrolase; zinc F:5-38/Domain: tetratricopeptide repeat homology <TRI>  
 F:39-72/Domain: tetratricopeptide repeat homology <TT2>  
 F:73-106/Domain: tetratricopeptide repeat homology <TT3>  
 F:179-444/Domain: phosphoprotein phosphatase core homology <PPP>  
 F:211-280/Domain: phosphoesterase core homology <PEC>  
 F:217,219,246/Binding site: iron (Asp, His, Asp) #status predicted  
 F:246,278,327,404/Binding site: zinc (Asp, Asn, His, His) #status predicted  
 F:249,279,428/Active site: Asp, His, Tyr #status predicted  
 F:250,377/Binding site: substrate phosphate (Arg) #status predicted

Query Match 47.6%; Score 39; DB 1; Length 473;  
 Best Local Similarity 66.7%; Pred. No. 63;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 ELGNEPNSFLKK 14  
 ||| :|||:  
 Db 7 ELKNEANKFLKE 18

RESULT 28  
 H84619  
 hypothetical protein Atg23050 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
 C:Accession: H84619  
 R:Liu, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.T.; Town, C.D.; Fujii, C.Y.; M M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
 A:Reference number: AB4420; MUID:20083487; PMID:10617197

A:Accession: H94619  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-481 <STO>

A:Cross-references: UNIPROT:O64814; UNIPARC:UPI000000C6D0; GB:AE002093; NID:g3169178; PID:G3169178  
 C:Genetics:

A:Gene: At2g23050  
 A:Map position: 2  
 C:Superfamily: Arabidopsis hypothetical protein F19F18.80

Query Match 47.6%; Score 39; DB 2; Length 481;  
 Best Local Similarity 58.3%; Pred. No. 65;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 ELGNPNFLKK 14

Db 5 KLGKPDSEFLSK 16

RESULT 29

S67621

hypothetical protein YDL085w - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein D2436

C:Species: *Saccharomyces cerevisiae*

C:Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 09-Jul-2004

C:Accession: S67621

R:Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67608

A:Accession: S67621

A:Molecule type: DNA

A:Residues: 1-545 <WAM>

A:Cross-references: UNIPROT:Q07500; UNIPARC:UPI000006C394; EMBL:Z74133; NID:g1431109; PID:g1431109

A:Experimental source: strain S288C

C:Genetics:

A:Gene: MIPS:YDL085w

A:Cross-references: SGD:S0002243

A:Map position: 4L

C:Superfamily: NADH dehydrogenase

Query Match 47.6%; Score 39; DB 2; Length 545;  
 Best Local Similarity 58.3%; Pred. No. 75;  
 Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 WELGNPNFLK 13

Db 364 WATGNEPDLFSK 375

RESULT 30

H95922

hypothetical membrane-anchored protein [imported] - *Sinorhizobium meliloti* (strain 1021)

C:Species: *Sinorhizobium meliloti*

C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 09-Jul-2004

C:Accession: H95922

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo

A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: H95922

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-554 <KUR>

A:Cross-references: UNIPROT:Q92VQ1; UNIPARC:UPI00000CB5E8; GB:AL591985; PIDN:CAC49048.1;

A:Experimental source: strain 1021, megaplasmid pSymB

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, J.  
 hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: Smb21069

A:Genome: plasmid

Query Match 47.6%; Score 39; DB 2; Length 554;

Best Local Similarity 60.0%; Pred. No. 76;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 WELGNPNFSF 11

Db 205 WELGNEMENY 214

RESULT 31

AD2408

hypothetical protein all4820 [imported] - *Nostoc* sp. (strain PCC 7120)

C:Species: *Nostoc* sp. PCC 7120

A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004

C:Accession: AD2408

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,

Nakazaki, N.; Shimoto, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena*

A:Reference number: AB1807; MUID:21595285; PMID:11759840

C:Accession: AD2408

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-628 <KUR>

A:Cross-references: UNIPROT:O8YMW6; UNIPARC:UPI000000C6C70; GB:BA000019; PIDN:BA076519.1;

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all4820

Query Match 47.6%; Score 39; DB 2; Length 628;

Best Local Similarity 85.7%; Pred. No. 88;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 WELGNP 8

Db 419 WELGNRP 425

RESULT 32

AD3057

glycogen debranching enzyme glgX [imported] - *Agrobacterium tumefaciens* (strain C58, Dupont)

C:Species: *Agrobacterium tumefaciens*

C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 31-Dec-2004

C:Accession: AD3057

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;

erage, G.; Gallet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClelland,

Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AD3057

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-651 <KUR>

A:Cross-references: UNIPROT:Q8U8L7; UNIPARC:UPI00000D242A; GB:AE008689; PIDN:AAL44874.1;

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: glgX

A:Map position: linear chromosome

C:Superfamily: isoamylase-type debranching enzyme

```

Query Match          47.6%; Score 39; DB 2; Length 651;
Best Local Similarity 58.3%; Pred. No. 92;
Matches              7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      2 WELGNEPNSFLK 13
       :|||||:
Db      371 YOLGNFPFESFLE 382

RESULT 33
B98229
glycogen debranching enzyme (AJ291603) [imported] - Agrobacterium tumefaciens (strain C5
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 31-Dec-2004
C;Accession: B98229
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: B98229
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-651 <KUR>
A;Cross-references: UNIPROT:Q8U8L7; UNIPARC:UPI00000D242A; GB:AE007870; PIDN:AAK89356.1;
C;Genetics:
A;Gene: AGR L 1566
A;Map position: linear chromosome
C;Superfamily: isoamylase-type debranching enzyme

Query Match          47.6%; Score 39; DB 2; Length 651;
Best Local Similarity 58.3%; Pred. No. 92;
Matches              7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      2 WELGNEPNSFLK 13
       :|||||:
Db      371 YOLGNFPFESFLE 382

RESULT 34
I51684
epithelial sodium channel, gamma subunit - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C;Accession: I51684
R;Puoti, A.; May, A.; Canessa, C.M.; Horisberger, J.
Am. J. Physiol. 269, 188-197, 1995
A;Title: The highly selective, low conductance epithelial sodium channel of Xenopus lae
A;Reference number: I51682
A;Accession: I51684
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-660 <PUO>
A;Cross-references: UNIPROT:PS1171; UNIPARC:UPI0000135621; EMBL:U25342; NID:g886045; PID
C;Genetics:
A;Gene: GammaENaC
C;Superfamily: human amiloride-sensitive sodium channel protein; fibronectin type I repe
F;375-410/Domain: fibronectin type I repeat homology <IFR>

Query Match          47.6%; Score 39; DB 2; Length 660;
Best Local Similarity 57.1%; Pred. No. 93;
Matches              8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      1 SWELGNEPNSFLK 14
       :|||||:
Db      488 SWELGEXLNKLTG 501

RESULT 35
E72278
endo-1,4-beta-mannosidase - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004

```

```

C;Accession: D72278
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: D72278
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-669 <ARN>
A;Cross-references: UNIPROT:Q9XOV4; UNIPARC:UPI00000D38F2; GB:AE001779; GB:AE000512; NID:
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM1227

Query Match          47.6%; Score 39; DB 2; Length 669;
Best Local Similarity 75.0%; Pred. No. 94;
Matches              6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 SWELGNEP 8
       :|||||:
Db      193 AWELANEP 200

RESULT 36
S29043
cellulase (EC 3.2.1.4) - Bacillus sp.
N;Alternate names: endo-1,4-beta-glucanase
C;Species: Bacillus sp.
C;Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C;Accession: S29043; PC4404
R;Ozaki, K.; Shikata, S.; Kawai, S.; Ito, S.; Okamoto, K.
J. Gen. Microbiol. 136, 1327-1334, 1990
A;Title: Molecular cloning and nucleotide sequence of a gene for alkaline cellulase from
A;Reference number: S29043; MUID:91037937; PMID:2230718
A;Accession: S29043
A;Molecule type: DNA
A;Residues: 1-941 <OZA>
A;Cross-references: UNIPROT:P19424; UNIPARC:UPI000012BDF8; EMBL:M27420; NID:g142664; PIDN
R;Shirai, T.; Yamane, T.; Hida, T.; Kuyama, K.; Suzuki, A.; Ashida, T.; Ozaki, K.; Ito,
J. Biochem. 122, 683-685, 1997
A;Title: Crystallization and preliminary X-ray analysis of a truncated family A alkaline
A;Reference number: PC4404; MUID:98060488; PMID:9399567
A;Accession: PC4404
A;Molecule type: protein
A;Residues: 228-584 <SHI>
A;Cross-references: UNIPARC:UPI00001791AD
A;Experimental source: strain KSM-635
C;Function:
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel
A;Pathway: cellulose degradation
C;Superfamily: Bacillus sp. KSM-635 alkaline cellulase; S-layer repeat homology; Thermot
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;41-95/Domain: S-layer repeat homology <SLR1>
F;101-153/Domain: S-layer repeat homology <SLR2>
F;164-219/Domain: S-layer repeat homology <SLR3>
F;766-908/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA>

Query Match          47.6%; Score 39; DB 2; Length 941;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches              6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      2 WELGNEPN 9
       :|||||:
Db      368 WELANEPS 375

RESULT 37
E72215
oligopeptide ABC transporter, periplasmic oligopeptide-binding protein - Thermotoga marit
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004

```

C;Accession: E72215  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.  
C.M.  
Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A;Reference number: A72200; PMID:99287316; PMID:10360571  
A;Accession: E72215  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-642 <ARN>  
A;Cross-references: UNIPROT:Q9X268; UNIPARC:UPI00000D3860; GB:AE001813; GB:AE000512; NID  
A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TM1746

Query Match 47.0%; Score 38.5; DB 2; Length 642;  
Best Local Similarity 44.4%; Pred. No. 1.1e+02;  
Matches 8; Conservative 3; Mismatches 2; Indels 5; Gaps 1;  
2 WE-----LGNPNPNSFLKK 14  
||| : |||  
490 WEKQEVNSPDEFLKK 507

Db

RESULT 38  
G96948  
uncharacterized small conserved protein, homolog of YUKF/YFJA B. subtilis CAC0398 [import  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C;Accession: G96948  
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A;Reference number: A96900; PMID:21359325; PMID:21359325  
A;Accession: G96948  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-100 <KUR>  
A;Cross-references: UNIPROT:Q97M02; UNIPARC:UPI00000C9E5E; GB:AE001437; PIDN:AAK78378.1;  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC0398

Query Match 46.3%; Score 38; DB 2; Length 100;  
Best Local Similarity 54.5%; Pred. No. 16;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
1 SWELGNEPNSF 11  
||| | |  
73 AWELNNTATPF 83

Db

RESULT 39  
H87375  
hypothetical protein CC1020 [imported] - Caulobacter crescentus  
C;Species: Caulobacter crescentus  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
C;Accession: H87375  
R;Niernan, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolar  
N., J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A;Title: Complete Genome Sequence of Caulobacter crescentus.  
A;Reference number: A87249; PMID:21173698; PMID:11259647  
A;Accession: H87375  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-129 <STO>  
A;Cross-references: UNIPROT:Q9A9G8; UNIPARC:UPI00000C7244; GB:AE005673; NID:gl13422312; F  
C;Genetics:  
A;Gene: CC1020

Query Match 46.3%; Score 38; DB 2; Length 129;  
Best Local Similarity 42.9%; Pred. No. 22;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKK 14  
: ||| | |  
Db 108 AMLVGSEPKSLKR 121

RESULT 40  
F86239  
protein F20B24.4 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C;Accession: F86239  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ket, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; PMID:21016719; PMID:11130712  
A;Accession: F86239  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-135 <STO>  
A;Cross-references: UNIPROT:Q9SGV9; UNIPARC:UPI000009D28C; GB:AE005172; NID:G6573734; P  
C;Genetics:  
A;Gene: F20B24.4  
A;Map position: 1

Query Match 46.3%; Score 38; DB 2; Length 135;  
Best Local Similarity 50.0%; Pred. No. 23;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 WEELGNEPNSF 11  
||| | |  
Db 29 WKLGDDPSRF 38

RESULT 41  
AH1586  
bacteriophage phi-105 ORF2 protein homolog lin1233 [imported] - Listeria innocua (strain  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C;Accession: AH1586  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; PMID:21537279; PMID:11679669  
A;Accession: AH1586  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-139 <GLA>  
A;Cross-references: UNIPROT:Q92CD9; UNIPARC:UPI00000CC4E3; GB:AL592022; PIDN:CAC96464.1;  
A;Experimental source: strain Clp11262  
C;Genetics:  
A;Gene: lin1233  
C;Superfamily: phage phi-105 immunity region protein 2

Query Match 46.3%; Score 38; DB 2; Length 139;  
Best Local Similarity 50.0%; Pred. No. 24;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLK 13  
| | | | |  
Db 33 WNLANDTNGFYK 44

RESULT 42  
822788  
thymidylate kinase XF0590 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 31-Dec-2004  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A>Note: for a complete list of authors see reference number A59328 below  
A:Accession: F82788  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-217 <SIM>  
A:Cross-references: UNIPARC:UPI0000165A56; GB:AE003904; GB:AE003849; NID:g9105433; PIDN:  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.U.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tshuko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF0580  
C:Superfamily: thymidylate kinase

Query Match 46.3%; Score 38; DB 2; Length 217;  
Best Local Similarity 60.0%; Pred. No. 39;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WELGNEPNSF 11  
| | | | |  
Db 146 WERGSTEHPF 155

RESULT 43  
AE2956  
haloacid dehalogenase-like hydrolase [imported] - Agrobacterium tumefaciens (strain C58,  
C:Species: Agrobacterium tumefaciens  
C>Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 05-Oct-2004  
C:Accession: AE2956  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AE2956  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-228 <KUR>  
A:Cross-references: UNIPROT:Q8UAW6; UNIPARC:UPI00000D212D; GB:AE008689; PIDN:AAL44067.1;  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu3251  
A:Map position: linear chromosome  
C:Superfamily: Alcaligenes eutrophus phosphoglycolate phosphatase

Query Match 46.3%; Score 38; DB 2; Length 228;  
Best Local Similarity 46.7%; Pred. No. 42;  
Matches 7; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

QY 2 WELGN--EPNSFLKK 14  
| | | | |  
Db 41 WAVANGIEPNFLQR 55

## RESULT 44

A98327  
hypothetical protein AGR\_L\_3128 [imported] - Agrobacterium tumefaciens (strain C58, Cerc  
C:Species: Agrobacterium tumefaciens  
C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 05-Oct-2004  
C:Accession: A98327  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman, B.;  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; MUID:21608551; PMID:11743194  
A:Accession: A98327  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-228 <KUR>  
A:Cross-references: UNIPROT:Q8UAW6; UNIPARC:UPI00000D212D; GB:AE007870; PIDN:AAK90139.1;  
C:Genetics:  
A:Gene: AGR\_L\_3128  
A:Map position: linear chromosome  
C:Superfamily: Alcaligenes eutrophus phosphoglycolate phosphatase

Query Match 46.3%; Score 38; DB 2; Length 228;  
Best Local Similarity 46.7%; Pred. No. 42;  
Matches 7; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

QY 2 WELGN--EPNSFLKK 14  
| | | | |  
Db 41 WAVANGIEPNFLQR 55

## RESULT 45

S49780  
hypothetical protein YDR183w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein YD9395.17  
C:Species: Saccharomyces cerevisiae  
C>Date: 13-Jan-1995 #sequence\_revision 10-Feb-1995 #text\_change 09-Jul-2004  
C:Accession: S49780  
R:Murphy, L.; Harris, D.E.  
submitted to the EMBL Data Library, November 1994  
A:Reference number: S49764  
A:Accession: S49780  
A:Molecule type: DNA  
A:Residues: 1-230 <MUR>  
A:Cross-references: UNIPROT:Q04004; UNIPARC:UPI000013A8BF; EMBL:Z46727; NID:g1289283; PI  
C:Genetics:  
A:Gene: SGD:PLP1; MIPS:YDR183w  
A:Cross-references: SGD:S0002591  
A:Map position: 4R

Query Match 46.3%; Score 38; DB 2; Length 230;  
Best Local Similarity 66.7%; Pred. No. 42;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 ELGNEPNSF 11  
| | | | |  
Db 174 KLGNDFNGF 182

## RESULT 46

F70710  
probable 3-oxoacyl-[acyl-carrier protein] reductase - Mycobacterium tuberculosis (strain  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 05-Oct-2004

C;Accession: F70710  
R;Cole, S.T.; Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: F70710  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-247 <COL>  
A;Cross-references: UNIPROT:Q48930; UNIPARC:UPI000012A460; GB:Z79701; GB:AL123456; NID:9  
A;Experimental source: strain H37RV  
C;Genetics:  
A;Gene: fabI  
C;Superfamily: short-chain dehydrogenase; short-chain alcohol dehydrogenase homology  
F;16-184/Domain: short-chain alcohol dehydrogenase homology <SADH>  
Query Match 46.3%; Score 38; DB 2; Length 247;  
Best Local Similarity 46.7%; Pred. No. 46;  
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
QY 1 SWELGNEPNSFLKKA 15  
||| :||| :  
Db 144 SWGIGNQNYAASKA 158  
RESULT 47  
T38320  
hypothetical protein SPAC23H4.08 - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C;Accession: T38320  
R;Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V. submitted to the EMBL Data Library, September 1997  
A;Reference number: 221733  
A;Accession: T38320  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-277 <BRO>  
A;Cross-references: UNIPROT:O13951; UNIPARC:UPI000006AEOB; EMBL:Z98977; PIDN:CAB11662.1;  
A;Experimental source: strain 972h-; cosmid C23H4  
C;Genetics:  
A;Gene: SPDB:SPAC23H4.08  
A;Map position: 1  
A;Introns: 21/1; 44/1; 148/2  
Query Match 46.3%; Score 38; DB 2; Length 277;  
Best Local Similarity 87.5%; Pred. No. 52;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 ELGNPNS 10  
||| :||| :  
Db 23 ELGNPNS 30  
RESULT 48  
T24356  
hypothetical protein T02D1.3 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T24356  
R;Basham, V. submitted to the EMBL Data Library, December 1996  
A;Reference number: Z19880  
A;Accession: T24356  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-322 <WIL>  
A;Cross-references: UNIPROT:O45729; UNIPARC:UPI0000061119; EMBL:Z83319; PIDN:CAB05908.1;  
A;Experimental source: clone T02D1  
C;Genetics:

A;Gene: CESP:T02D1.3  
A;Map position: 4  
A;Introns: 77/3; 147/3; 200/3; 266/3  
C;Superfamily: Caenorhabditis elegans hypothetical protein C33A12.9b  
Query Match 46.3%; Score 38; DB 2; Length 322;  
Best Local Similarity 53.8%; Pred. No. 62;  
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
QY 1 SWELGNEPNSFLK 13  
||| :||| :  
Db 99 SWCASQKPNHFLK 111  
RESULT 49  
S09283  
fructose-bisphosphate aldolase (EC 4.1.2.13) - Corynebacterium glutamicum  
C;Species: Corynebacterium glutamicum  
C;Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 09-Jul-2004  
C;Accession: S09283  
R;von der Osten, C.H.; Barbas, C.F.; Wong, C.H.; Sinskey, A.J. Mol. Microbiol. 3, 1625-1637, 1989  
A;Title: Molecular cloning, nucleotide sequence and fine-structural analysis of the Coryne-  
ass I and class II aldolases.  
A;Reference number: S09283; MUID:90136092; PMID:2615658  
A;Accession: S09283  
A;Molecule type: DNA  
A;Residues: 1-344 <OST>  
A;Cross-references: UNIPROT:P19537; UNIPARC:UPI000016EAEA; EMBL:X17313; NID:G40494; PIDN:  
C;Genetics:  
A;Gene: fda  
C;Superfamily: fructose-bisphosphate aldolase II  
C;Keywords: aldehyde-lyase; carbon-carbon lyase; gluconeogenesis; glycolysis  
Query Match 46.3%; Score 38; DB 2; Length 344;  
Best Local Similarity 47.1%; Pred. No. 67;  
Matches 8; Conservative 4; Mismatches 1; Indels 4; Gaps 1;  
QY 3 ELGN----EPNSFLKKA 15  
||| :||| :  
Db 304 EVGNKKAYDPSYMKKA 320  
RESULT 50  
T04762  
chitinase homolog T16H5.170 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C;Accession: T04762  
R;Bevan, M.; De Haan, M.; Maarse, A.C.; Grivell, L.A.; Bancroft, I.; Mewes, H.W.; Mayer, submitted to the Protein Sequence Database, June 1998  
A;Reference number: Z15383  
A;Accession: T04762  
A;Molecule type: DNA  
A;Residues: 1-379 <BEV>  
A;Cross-references: UNIPROT:O81862; UNIPARC:UPI000009E965; EMBL:AL024486  
A;Experimental source: cultivar Columbia; BAC clone T16H5  
C;Genetics:  
A;Map position: 4  
A;Introns: 232/1; 362/1  
A;Note: T16H5.170  
C;Superfamily: Streptomyces chitinase chi40  
Query Match 46.3%; Score 38; DB 2; Length 379;  
Best Local Similarity 46.7%; Pred. No. 75;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
QY 1 SWELGNEPNSFLKKA 15  
||| :||| :  
Db 347 SWHVGADDNSGLSRA 361  
Search completed: June 5, 2006, 12:53:44



Job time : 32.6438 secs

---

OM protein - protein search, using sw model  
 Run on: June 5, 2006, 12:32:17 ; Search time 131.507 Seconds  
 (without alignments)  
 105.510 Million cell updates/sec

Title: US-10-645-659A-8  
 Perfect score: 82  
 Sequence: 1 SWELGNPNFLKKA 15

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 1000 summaries

Database : UniProt 7.2.1\*  
 1: uniprot\_sprot;\*  
 2: uniprot\_trembl;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	100.0	543	1 HPSE HUMAN	Q9Y251 homo sapien
2	76	92.7	535	1 HPSE MOUSE	Q6Ygz1 mus musculus
3	76	92.7	536	1 HPSE RAT	Q71rp1 rattus norv
4	76	92.7	558	2 Q33X5 SPJJD	Q333x5 spalax juda
5	76	92.7	574	2 Q33X6 SPJJD	Q333x6 spalax juda
6	76	92.7	574	2 Q33X7_9RODE	Q333x7 spalax carm
7	76	92.7	574	2 Q33X8_9RODE	Q333x8 spalax gola
8	76	92.7	574	2 Q33X9_9RODE	Q333x9 spalax gali
9	73	89.0	523	1 HPSE CHICK	Q90yk5 gallus gall
10	73	89.0	545	1 HPSE BOVIN	Q9myv0 bos taurus
11	72	87.8	255	2 Q4TGB TTING	Q4TGB8 tetraodon n
12	72	87.8	533	2 Q4SYF TTING	Q4SYF6 tetraodon n
13	58	70.7	592	1 HPSE2 HUMAN	Q8wwq2 homo sapien
14	58	70.7	592	2 Q2M1H5 HUMAN	Q2m1h9 homo sapien
15	58	70.7	597	2 Q4TB80 TTING	Q4tb80 tetraodon n
16	57	69.5	515	2 Q8T108 BOMMO	Q8t108 bombyx mori
17	52	63.4	651	1 BGLR CANFA	O18835 canis famil
18	51	62.2	652	2 Q4PAT7 FIG	Q4pat7 sus scrofa
19	50	61.0	651	1 BGLR FELCA	Q97524 felis silve
20	49	59.8	581	2 Q6FYV6 BARQU	Q6fyv6 bartonella
21	48	58.5	260	2 Q2TYD4 ASPOR	Q2tyd4 aspergillus
22	48	58.5	435	2 Q4Y35_BACFR	Q64y35 bacteroides
23	47	57.3	1297	2 Q31KG9 PSEHT	Q31kg9 pseudosalter
24	46	56.1	627	2 Q3CKA3 THEET	Q3cka3 thermoanaer
25	46	56.1	637	2 Q8RDH1 THETN	Q8rdh1 thermoanaer
26	46	56.1	651	2 Q3E4N3 CHLAU	Q3e4n3 chloroflexu
27	45	54.9	373	2 Q2U2I3 ASPOR	Q2u2i3 aspergillus
28	45	54.9	409	2 Q5B833 EMENI	Q5b833 aspergillus
29	45	54.9	437	2 Q99036 TRIRE	Q99036 trichoderma
30	45	54.9	490	2 Q3DZN2 CHLAU	Q3dzn2 chloroflexu
31	45	54.9	538	2 Q5CW71_CRYPV	Q5cw71 cryptospori

32	45	54.9	648	1 BGLR MOUSE	P12265 mus musculus
33	45	54.9	648	2 Q3TW82 MOUSE	Q3tw82 m osteoclas
34	45	54.9	648	2 Q6IR10_MOUSE	Q6ir10 m glucuroni
35	45	54.9	1687	2 Q4IR74_GIBZE	Q4ir74 gibberella
36	44	53.7	90	2 Q3D554 STRAG	Q3d554 streptococc
37	44	53.7	90	2 Q3DNV8 STRAG	Q3dnv8 streptococc
38	44	53.7	90	2 Q48S82 STRPM	Q48s82 streptococc
39	44	53.7	90	2 Q8DX45_STRAS	Q8dx45 streptococc
40	44	53.7	144	2 Q4G325_MACMU	Q4g325 macaca mula
41	44	53.7	379	1 TSYL1_MOUSE	Q8852 mus musculus
42	44	53.7	379	2 Q3TKW0_MOUSE	Q3tkw0 m blastocys
43	44	53.7	379	2 Q642B1_RAT	Q642b1 rattus norv
44	44	53.7	437	2 Q5TFE6_HUMAN	Q5tf6 homo sapien
45	44	53.7	438	1 TSYL1_HUMAN	Q9h09 homo sapien
46	44	53.7	438	2 Q6FI91_HUMAN	Q6fi91 homo sapien
47	44	53.7	438	2 Q5R5G8_PONPY	Q5r5g8 pongo pygma
48	44	53.7	438	2 Q5RA88_PONPY	Q5ra88 pongo pygma
49	44	53.7	497	2 Q5QNQ4_MOUSE	Q5gnq4 mus musculu
50	43.5	53.0	492	2 Q47BM1_DECAR	Q47bm1 dechloromon
51	43	52.4	75	2 Q4INV6_METBU	Q4inv6 methanococc
52	43	52.4	170	2 Q5LQE4_SILPO	Q5lqe4 sillicibacte
53	43	52.4	201	2 Q3EJL4_BACTI	Q3ejl4 bacillus th
54	43	52.4	310	2 Q5SBQ2_DICDI	Q5sbq2 dictyosteli
55	43	52.4	321	2 Q3L5C1_BIFBR	Q3l5c1 bifidobacte
56	43	52.4	321	2 Q8G872_BIFLO	Q8g872 bifidobacte
57	43	52.4	365	2 Q4AYT0_9BURK	Q4ayt0 polaromonas
58	43	52.4	450	2 Q6ISB8_CABER	Q6isb8 caenorhabdi
59	43	52.4	623	2 Q7RY27_NEUCR	Q7ry27 neurospora
60	43	52.4	624	2 Q9PEA0_NEUCR	Q9pea0 neurospora
61	43	52.4	701	2 Q600U1_MYCH2	Q600u1 mycoplasma
62	43	52.4	717	2 Q4A816_MYCH7	Q4a816 mycoplasma
63	43	52.4	717	2 Q4A9Y6_MYCHJ	Q4a9y6 mycoplasma
64	43	52.4	1080	2 Q9CM43_PASMU	Q9cm43 pasteurella
65	43	52.4	2288	2 Q5AXQ5_EMENI	Q5axq5 aspergillus
66	42.5	51.8	166	2 Q4H5P7_9DEIO	Q4h5p7 deinococcus
67	42	51.2	59	2 Q3YOH0_ENTFC	Q3yoh0 enterococcu
68	42	51.2	88	2 Q7TQ6_PRONM	Q7ttq6 prochloroco
69	42	51.2	110	2 Q87GD2_VIBPA	Q87gd2 vibrio para
70	42	51.2	127	2 Q39XW1_GBOMG	Q39xw1 geobacter m
71	42	51.2	139	1 CYC22_RHOPA	P00091 rhodopsedo
72	42	51.2	139	2 Q8GI80_RHOPA	Q8gi80 rhodopsedo
73	42	51.2	163	2 Q8RME6_RHOPA	Q8rme6 rhodopsedo
74	42	51.2	250	1 THYX_CORGL	P40111 corynebacte
75	42	51.2	321	2 Q61YY6_CABER	Q61yy6 caenorhabdi
76	42	51.2	416	2 Q2SQ73_9GAMM	Q2sq73 hahella che
77	42	51.2	436	2 Q421D4_DESHA	Q421d4 desulfitoba
78	42	51.2	489	2 Q4WH6_GIBZE	Q4wh6 gibberella
79	42	51.2	552	2 Q4IBY1_GIBZE	Q4iby1 gibberella
80	42	51.2	558	2 Q4CFT4_CLOPM	Q4cft4 clostridium
81	42	51.2	630	2 Q6CT20_KLULA	Q6ct20 kluyveromyc
82	42	51.2	758	2 Q5AF48_CANAL	Q5af48 candida alb
83	42	51.2	779	2 Q3FSA5_9BURK	Q3fsa5 burkholderi
84	42	51.2	779	2 Q4BIB6_BURVI	Q4bib6 burkholderi
85	42	51.2	904	2 Q87JH7_VIBPA	Q87jh7 vibrio para
86	42	51.2	947	1 YBV6_YEAST	P38286 saccharomyc
87	42	51.2	947	2 Q2SSN6_9SPHI	Q2ssn6 salinibacte
88	42	51.2	1420	2 Q5AFT3_CANAL	Q5aft3 candida alb
89	42	51.2	4760	2 Q3VV55_PROAE	Q3vv55 prosthecocb
90	41.5	50.6	118	2 Q81CK7_BACCR	Q81ck7 bacillus ce
91	41.5	50.6	352	1 XYNA_DCTCH	Q12603 dictyoglomu
92	41	50.0	48	2 Q5G5M1_9HIV1	Q5g5m1 human immun
93	41	50.0	81	2 Q3SDLO_NEUCR	Q3sdlo neurospora
94	41	50.0	87	2 Q3HP06_9HIV1	Q3hp06 human immun
95	41	50.0	97	2 Q4A8N6_MYCH7	Q4a8n6 mycoplasma
96	41	50.0	97	2 Q4AAK5_MYCHJ	Q4aa5 mycoplasma
97	41	50.0	121	2 Q5X4L3_LEGPA	Q5x4l3 legionella
98	41	50.0	121	2 Q3UX22_MOUSE	Q3ux22 mus musculu
99	41	50.0	123	2 Q5WM03_LEGPL	Q5wm03 legionella
100	41	50.0	123	2 Q5ZUU4_LEGPH	Q5zuu4 legionella
101	41	50.0	127	2 Q8IHU0_PLAF7	Q8ihu0 plasmodium
102	41	50.0	158	2 Q6NJF3_CORDI	Q6njf3 corynebacte
103	41	50.0	163	2 Q5OU98_ENTHI	Q5ou98 entamoeba h
104	41	50.0	172	2 Q4HQ21_CAMUP	Q4hq21 campylobact

105	41	50.0	176	2	Q32LJ0_SHIBS	Q32lj0 shigella bo	178	40	48.8	431	2	05KLX9_CRYNE	Q5klx9 cryptococcus
106	41	50.0	181	2	Q8MU93_ANCCA	Q8mu93 ancylostoma	179	40	48.8	445	2	Q7N7C7_PHOLL	Q7n7c7 phorhabdu
107	41	50.0	194	2	Q9HU69_PSEAE	Q9hu69 pseudomonas	180	40	48.8	478	2	Q8SW72_ENCCU	Q8sw72 entcephalito
108	41	50.0	201	2	Q6HLQ1_BACILLUS TH	Q6hlq1 bacillus th	181	40	48.8	498	2	Q9DH06_9HIV1	Q9dh06 human immun
109	41	50.0	217	2	Q3A6Z5_PBLCKD	Q3a6z5 pelobacter	182	40	48.8	498	2	Q9QMM1_9HIV1	Q9qmm1 human immun
110	41	50.0	274	1	Y059_CAEBL	Q09465 caenorhabdi	183	40	48.8	500	2	Q75W19_PANGI	Q75w19 panax ginse
111	41	50.0	281	2	Q5KH72_MYCMO	Q6khy2 mycoplasma	184	40	48.8	502	2	Q9VFS1_DROME	Q9vfs1 drosophila
112	41	50.0	298	1	ABILI_ATH	Q88ms arabidopsis	185	40	48.8	505	1	TCMO_ARATH	P92994 arabidopsis
113	41	50.0	303	2	Q9ND56_DROME	Q9nd56 drosophila	186	40	48.8	505	2	Q9T0N4_ARATH	Q9t0n4 arabidopsis
114	41	50.0	303	2	Q9VFG5_DROME	Q9vfg5 drosophila	187	40	48.8	508	2	Q4S6P2_TETNG	Q4s6p2 tetraodon n
115	41	50.0	328	2	Q4V200_BACCZ	Q4v200 bacillus ce	188	40	48.8	512	2	Q3RKL2_RALME	Q3rkl2 raleconia m
116	41	50.0	335	2	Q61YV5_CAEBR	Q61yvs caenorhabdi	189	40	48.8	551	2	Q61RL8_XENLA	Q61rl8 xenopus lae
117	41	50.0	429	2	Q733G5_BACC1	Q733gs bacillus ce	190	40	48.8	559	2	Q7SFB0_NEUCR	Q7sfb0 neurospora
118	41	50.0	447	2	Q9X1I8_ARATH	Q9x1i8 arabidopsis	191	40	48.8	611	2	Q3E127_CHLAU	Q3e127 chloroflexu
119	41	50.0	452	2	Q62KG8_BURMA	Q62kg8 burkholderi	192	40	48.8	621	2	Q3EMW3_BACT1	Q3emw3 bacillus th
120	41	50.0	461	2	Q47YX8_COLP3	Q47yx8 colwellia p	193	40	48.8	628	2	Q95Q32_CAEEL	Q95q32 caenorhabdi
121	41	50.0	510	2	Q3JT27_BURP1	Q3jct27 burkholderi	194	40	48.8	640	2	Q9K0R9_NEIMB	Q9k0r9 neisseria m
122	41	50.0	510	2	Q63T11_BURPS	Q63t11 burkholderi	195	40	48.8	642	2	Q3UVE7_MOUSE	Q3uve7 mus musculus
123	41	50.0	514	2	Q87KV6_VIBPA	Q87kv6 vibrio para	196	40	48.8	648	1	BGLR_CERAE	O77695 cercopitheci
124	41	50.0	529	2	Q6ZJE2_ORISA	Q6zje2 oryza sativ	197	40	48.8	651	1	BGLR_HUMAN	P08236 homo sapien
125	41	50.0	533	2	Q973V9_SULTO	Q973v9 sulfolobus	198	40	48.8	651	2	Q549U0_HUMAN	Q549u0 homo sapien
126	41	50.0	533	2	Q7NSM0_PHOLL	Q7nsm0 photorhabdu	199	40	48.8	651	2	Q5R5N6_PONPY	Q5r5n6 pongo pygma
127	41	50.0	553	2	Q43N00_SOLUS	Q43nq0 solibacter	200	40	48.8	651	2	Q5R8A1_PONPY	Q5r8a1 pongo pygma
128	41	50.0	558	2	Q2SV40_BURTH	Q2svc4 burkholderi	201	40	48.8	662	2	Q92M11_RHIME	Q92m11 rhizobium m
129	41	50.0	632	2	Q4RJ08_TETNG	Q4rjq8 tetraodon n	202	40	48.8	694	2	Q9Z1R7_BACST	Q9z1r7 bacillus st
130	41	50.0	648	1	BGLR_RAT	P06760 rattus norv	203	40	48.8	716	1	IF2_BACSU	Q65ji1 bacillus li
131	41	50.0	650	2	Q4BZ00_CROWT	Q4bz00 crocosphaer	204	40	48.8	716	2	Q65J11_BACLD	P17889 bacillus su
132	41	50.0	734	2	Q7QDL1_ANOGA	Q7qdl1 anopheles g	205	40	48.8	799	2	Q5AGB1_CANAL	Q5agb1 candida alb
133	41	50.0	808	2	Q7TRJ3_RAT	Q7trpj3 rattus norv	206	40	48.8	820	2	Q82QC4_STPAW	Q82qc4 streptomyc
134	41	50.0	849	2	Q8WZQ5_NEUCR	Q8wzq5 neurospora	207	40	48.8	870	2	Q6FJW0_CANGA	Q6fjw0 candida gla
135	41	50.0	999	2	Q87B86_NEUCR	Q87b86 neurospora	208	40	48.8	879	2	Q4E4X9_TRYCR	Q4e4x9 trypanosoma
136	41	50.0	1148	2	Q918V6_CHICK	Q918v6 gallus gall	209	40	48.8	884	2	Q50UK8_ENTHI	Q50uk8 entamoeba h
137	41	50.0	1148	2	Q91BA7_CHICK	Q91ba7 gallus gall	210	40	48.8	950	2	Q6ZQE0_MOUSE	Q6zqe0 mus musculus
138	41	50.0	1182	2	Q5PQJ5_RAT	Q5pqj5 rattus norv	211	40	48.8	969	2	Q6BKM3_DEBHA	Q6bkm3 debaryomyce
139	41	50.0	1244	2	Q4P262_USTMA	Q4p262 uestilago ma	212	40	48.8	1058	1	U202A_ARATH	Q9x1k4 arabidopsis
140	41	50.0	2100	2	Q81Y32_BACAN	Q8ly32 bacillus an	213	40	48.8	1091	2	Q4N5Q3_THEPA	Q4n5q3 theileria p
141	41	50.0	2113	2	Q6HVD0_BACAN	Q6hvd0 bacillus an	214	40	48.8	1641	1	Y518_XYLFT	Q9pdx7 xyella fas
142	41	50.0	2213	2	Q8IHW8_PLAF7	Q8lhw8 plasmodium	215	40	48.8	1641	1	Y518_XYLFT	Q87e11 xyella fas
143	41	50.0	2520	2	Q4MMN6_BACC2	Q4mm6 bacillus ce	216	40	48.8	1641	2	Q3R2N7_XYLFA	Q3r2n7 xyella fas
144	41	50.0	2520	2	Q637G8_BACCZ	Q637g8 bacillus th	217	40	48.8	1641	2	Q3R898_XYLFA	Q3r898 xyella fas
145	40.5	49.4	314	2	Q6HFE0_BACHK	Q6hfe0 bacillus th	218	40	48.8	1641	2	Q3RGS2_XYLFA	Q3rgs2 xyella fas
146	40.5	49.4	314	2	Q8IFV3_GIALA	Q8ifv3 giardia lam	219	40	48.8	1880	2	Q8S8U2_ATRBE	Q8s8u2 atropa bell
147	40.5	49.4	314	2	Q7R527_GIALA	Q7r527 giardia lam	220	40	48.8	2898	2	Q61ZC4_CAEBR	Q61zc4 caenorhabdi
148	40.5	49.4	345	2	Q72V05_LEPIC	Q72v05 leptospira	221	39.5	48.2	153	1	SODC_CANAL	Q59924 candida alb
149	40.5	49.4	355	2	Q8EZL1_LEPIN	Q8ezl1 leptospira	222	39.5	48.2	332	2	Q3GNC6_9GAMM	Q3gnc6 psychobact
150	40.5	48.8	48	2	Q5GSP9_HIV1	Q5gsp9 human immun	223	39.5	48.2	490	2	Q3Z9W6_DEHEL	Q3z9w6 dehalococco
151	40	48.8	86	2	Q3HNX0_9HIV1	Q3hnx0 human immun	224	39.5	48.2	491	2	Q3ZWS1_DEHSC	Q3zws1 dehalococco
152	40	48.8	110	2	Q72TQ2_LEPIC	Q72tq2 leptospira	225	39.5	48.2	567	2	Q84VA5_ORYSA	Q84va5 oryza sativ
153	40	48.8	117	2	Q8F117_LEPIN	Q8f117 leptospira	226	39.5	48.2	594	1	GLMN_HUMAN	Q92990 homo sapien
154	40	48.8	169	2	Q8UIM6_AGRRT5	Q8uim6 agrobacteri	227	39.5	48.2	594	2	Q5VVC3_HUMAN	Q5vvc3 homo sapien
155	40	48.8	193	2	Q3HJ56_TYRER	Q3hj56 trichodesmi	228	39.5	48.2	608	2	Q8I3C5_PLAF7	Q8i3c5 plasmodium
156	40	48.8	201	2	Q73BL2_BACC1	Q73bl2 bacillus ce	229	39.5	48.2	676	2	Q8EFLM_BACST	Q8eln4 bacillus st
157	40	48.8	205	2	Q813U5_BACCR	Q813u5 bacillus ce	230	39.5	48.2	1058	2	Q69PQ1_ORYSA	Q69pq1 oryza sativ
158	40	48.8	231	2	Q66PV6_9CNID	Q66pv6 acropora ac	231	39.5	48.2	1377	2	Q804X3_CHICK	Q804x3 gallus gall
159	40	48.8	243	2	Q5BM98_SCHJA	Q5bm98 schistosoma	232	39	47.6	77	2	Q5PAW7_ANAMM	Q5paw7 anaplasma m
160	40	48.8	270	2	Q3GZ14_CABEL	Q3gz14 caenorhabdi	233	39	47.6	94	1	UCRQ_YEAST	P08525 saccharomyc
161	40	48.8	275	2	Q8VZY1_MUSAC	Q8vzy1 musa acumin	234	39	47.6	113	2	Q6PYA4_9ARAC	Q6pya4 frabonattus
162	40	48.8	303	2	Q44J85_CHRSL	Q44j85 chromohalob	235	39	47.6	126	2	Q3W103_9ACTO	Q3w103 frankia sp.
163	40	48.8	305	2	Q98FT1_RHILO	Q98ft1 rhizobium l	236	39	47.6	196	2	Q8DMC9_SYNEL	Q8dmc9 synecococc
164	40	48.8	331	2	Q45152_BACPR	Q45152 bacteroides	237	39	47.6	199	2	Q8XHK3_DROYA	Q8xhk3 drosophila
165	40	48.8	337	2	Q4L9G9_STAHTJ	Q4l9g9 staphylococ	238	39	47.6	201	2	Q81CJ2_BACCR	Q81cj2 bacillus ce
166	40	48.8	338	2	Q7P7K9_FUSNV	Q7p7k9 fusobacteri	239	39	47.6	201	2	Q3ELN8_BACTI	Q3eln8 bacillus th
167	40	48.8	343	2	Q44KD6_CHRSL	Q44kd6 chromohalob	240	39	47.6	201	2	Q63E76_BACCZ	Q63e76 bacillus ce
168	40	48.8	344	2	Q4JXT9_CORJK	Q4jxt9 corynebacte	241	39	47.6	218	2	Q81TH1_BACAN	Q81th1 bacillus an
169	40	48.8	344	2	Q5YNTK_NOCFA	Q5yntk nocardia fa	242	39	47.6	232	2	Q7RTI0_PLAYO	Q7rti0 plasmodium
170	40	48.8	344	2	Q8FMA6_COREF	Q8fma6 corynebacte	243	39	47.6	232	2	Q8RGQ6_FUSNK	Q8rgq6 fusobacteri
171	40	48.8	357	2	Q7MWH4_PORGI	Q7mwh4 porphyronon	244	39	47.6	234	2	Q3FLV3_9BURN	Q3flv3 rhodofera
172	40	48.8	370	2	Q4XV87_PLACH	Q4xv87 plasmodium	245	39	47.6	243	2	Q41GG3_GIRZE	Q41gg3 gibberella
173	40	48.8	413	1	V5Y3_TRYCK	P20949 trypanosoma	246	39	47.6	244	2	Q8WVK8_HUMAN	Q8wvk8 homo sapien
174	40	48.8	419	2	Q3C1D9_9BURK	Q3c1d9 comamonas s	247	39	47.6	275	2	Q7P645_FUSNV	Q7p645 fusobacteri
175	40	48.8	421	2	Q3CNA4_ALTAT	Q3cna4 pseudalter	248	39	47.6	276	2	Q4C8W3_CROWT	Q4c8w3 crocosphaer
176	40	48.8	431	2	Q55Y11_CRYNE	Q55y11 cryptococcus	249	39	47.6	279	2	Q4WWW4_ASPPU	Q4www4 aspergillus
177	40	48.8					250	39	47.6				

251	39	47.6	280	2	Q5VX98 HUMAN	Q5vx98 homo sapien	324	39	47.6	663	1	SCNNH_XENLA	Q13263 xenopus lae
252	39	47.6	294	2	Q5K2L9_BACLI	Q5k2l9 bacillus li	325	39	47.6	663	2	Q6NR19_XENLA	Q6nr19 xenopus lae
253	39	47.6	295	2	Q7V3P3_PROMP	Q7v3p3 prochloroco	326	39	47.6	666	2	Q9RIK7_THENA	Q9rik7 thermotoga
254	39	47.6	298	2	Q7VB69_PROMA	Q7vb69 prochloroco	327	39	47.6	669	2	Q9X0V4_THENA	Q9x0v4 thermotoga
255	39	47.6	300	2	Q3UV48_MOUSE	Q3uv48 mus musculus	328	39	47.6	680	2	Q9RIK9_THENA	Q9rik9 thermotoga
256	39	47.6	303	2	Q5VX97_HUMAN	Q5vx97 homo sapien	329	39	47.6	691	2	Q89NA6_BRAJA	Q89na6 bradyrhizob
257	39	47.6	303	2	Q8DU06_STRMU	Q8du06 streptococ	330	39	47.6	699	2	Q59ST4_CANAL	Q59st4 candida alb
258	39	47.6	312	2	Q2JUX9_9CVAN	Q2jux9 cyanobacter	331	39	47.6	700	2	Q4CNA6_TRYCR	Q4cna6 trypanosoma
259	39	47.6	325	2	Q9HP97_HALSA	Q9hp97 halobacteri	332	39	47.6	706	2	Q3HA64_TRIER	Q3ha64 trichodesmi
260	39	47.6	325	2	Q20705_CAEEL	Q20705 caenorhabdi	333	39	47.6	719	2	Q4KCP5_PSEFS	Q4kcp5 pseudomonas
261	39	47.6	335	2	Q6QOU1_CABBR	Q6qou1 caenorhabdi	334	39	47.6	771	2	Q8BME3_KLUDE	Q8bme3 mus musculus
262	39	47.6	349	2	Q8RGW9_FUSNN	Q8rgw9 fusobacteri	335	39	47.6	833	2	Q87AN6_MOUDE	Q87an6 kluyveromyc
263	39	47.6	354	2	Q9VCT5_DROME	Q9vct5 drosophila	336	39	47.6	900	2	Q2UQF4_ASPOK	Q2uqf4 aspergillus
264	39	47.6	358	2	Q4IHQ7_GIBZE	Q4ihq7 gibberella	337	39	47.6	907	2	Q4XUP4_PLACH	Q4xup4 plasmodium
265	39	47.6	363	2	Q82I33_STRAM	Q82i33 streptomyce	338	39	47.6	925	2	Q9HLE8_THEAC	Q9hle8 thermoplasma
266	39	47.6	366	2	Q81863_ARATH	Q81863 arabidopsis	339	39	47.6	941	1	GUN_BACS6	P19424 bacillus sp
267	39	47.6	371	2	Q315I3_DESDG	Q315i3 desulfovibr	340	39	47.6	995	2	Q2JN52_9CVAN	Q2jns2 cyanobacter
268	39	47.6	372	2	Q2EN0_DRSVH	Q2en0 desulfovibr	341	39	47.6	1126	2	Q8TKN1_METAC	Q8tkn1 methanosaer
269	39	47.6	377	2	Q00012_ASPAC	Q00012 aspergillus	342	39	47.6	1197	2	Q54HF9_DICDI	Q54hf9 dictyosteli
270	39	47.6	381	2	Q5AR04_EMENI	Q5ar04 aspergillus	343	39	47.6	1807	2	Q4WFF3_ASPFU	Q4wff3 aspergillus
271	39	47.6	381	2	Q5AVP1_EMENI	Q5avp1 aspergillus	344	39	47.6	1887	2	Q2VEC3_SOLBU	Q2vec3 solanum tub
272	39	47.6	383	2	Q2LE69_9EURO	Q2le69 aspergillus	345	39	47.6	1887	2	Q2MIC9_SOLAN	Q2mic9 solanum bul
273	39	47.6	383	2	Q5B7X2_EMENI	Q5b7x2 aspergillus	346	39	47.6	1948	2	Q4ABK7_BRARP	Q4abk7 brassica ra
274	39	47.6	386	2	Q2TXJ2_ASPOR	Q2txj2 aspergillus	347	39	47.6	2206	2	Q5B7N7_EMENI	Q5b7n7 aspergillus
275	39	47.6	390	2	Q48NR5_PSE14	Q48nr5 pseudomonas	348	39	47.6	2823	2	Q7RCM8_PLAYO	Q7rgm8 plasmodium
276	39	47.6	390	2	Q4ZMJ2_PSEU2	Q4zmj2 pseudomonas	349	39	47.6	2853	2	Q4PH47_USTWA	Q4ph47 ustilago ma
277	39	47.6	395	2	Q5J16_BACLD	Q5j16 bacillus li	350	38.5	47.0	642	2	Q9X268_THEMA	Q9x268 thermotoga
278	39	47.6	399	2	Q5AZ53_EMENI	Q5az53 aspergillus	351	38	46.3	59	2	Q4XEV0_PLACH	Q4xev0 plasmodium
279	39	47.6	405	2	Q4WIY0_ASPFU	Q4wiy0 aspergillus	352	38	46.3	74	2	Q4YU74_PLABE	Q4yu74 plasmodium
280	39	47.6	407	2	Q4NAS4_9MCC	Q4nas4 arthrobacte	353	38	46.3	97	2	Q424A2_DESHA	Q424a2 desulfitoba
281	39	47.6	410	2	Q93X40_LACSA	Q93x40 lactuca eat	354	38	46.3	100	2	Q6ZEG7_SYNY3	Q6zeg7 synechocyst
282	39	47.6	420	2	Q2ZDNE_CALSA	Q2zrne caldicellul	355	38	46.3	100	2	Q97M02_CLOAB	Q97m02 clostridium
283	39	47.6	430	2	Q8KJN4_RHILO	Q8kjin4 rhizobium l	356	38	46.3	129	2	Q9A9G8_CAUCR	Q9a9g8 caulobacter
284	39	47.6	431	2	Q580A2_9TRYP	Q580a2 trypanosoma	357	38	46.3	132	2	Q3ATF2_CHLCR	Q3atf2 chlorobium
285	39	47.6	433	2	Q465Q9_METBA	Q465q9 methanosaer	358	38	46.3	135	2	Q9SGY9_ARATH	Q9sgy9 arabidopsis
286	39	47.6	438	2	Q660F4_BORGA	Q660f4 borrelia ga	359	38	46.3	136	2	Q4ALT6_9ZZZZ	Q4alt6 uncultured
287	39	47.6	439	2	Q92401_AGABI	Q92401 agarius bi	360	38	46.3	136	2	Q4ALT9_9ZZZZ	Q4alt9 uncultured
288	39	47.6	439	2	Q9P893_AGABI	Q9p893 agarius bi	361	38	46.3	139	2	Q92CD9_LISIN	Q92cd9 listeria in
289	39	47.6	440	2	Q3MX18_9DELT	Q3mx18 syntrophoba	362	38	46.3	142	2	Q4ALT2_9ZZZZ	Q4alt2 uncultured
290	39	47.6	442	2	Q989J0_RHILO	Q989j0 rhizobium l	363	38	46.3	144	2	Q4ALT0_9ZZZZ	Q4alt0 uncultured
291	39	47.6	442	2	Q9AA10_CAUCR	Q9aa10 caulobacter	364	38	46.3	147	2	Q4ALT5_9ZZZZ	Q4alt5 uncultured
292	39	47.6	456	2	Q6QT42_9GAMM	Q6qt42 cellvibrio	365	38	46.3	147	2	Q4ALT8_9ZZZZ	Q4alt8 uncultured
293	39	47.6	459	2	Q31661_BORBU	Q31661 borrelia bu	366	38	46.3	150	2	Q4UBK3_THEAN	Q4ubk3 theileria a
294	39	47.6	465	2	Q3F8S9_9BURK	Q3f8s9 burkholderi	367	38	46.3	155	2	Q4N3Y6_THEPA	Q4n3y6 theileria p
295	39	47.6	465	2	Q457X3_9BURK	Q457x3 burkholderi	368	38	46.3	159	2	Q5DN68_9CAUD	Q5dn68 bacterioph
296	39	47.6	465	2	Q4LMP8_9BURK	Q4lmp8 burkholderi	369	38	46.3	160	2	Q8CM71_STRAS	Q8cm71 streptococ
297	39	47.6	465	2	Q39EV8_BURS3	Q39ev8 burkholderi	370	38	46.3	164	2	Q3AYM2_SYNS9	Q3aym2 synechococ
298	39	47.6	468	2	Q3X0X6_9ACTN	Q3x0x6 rubrobacter	371	38	46.3	166	2	Q4BWG7_CROWT	Q4bwg7 crocospaer
299	39	47.6	473	2	Q43049_SCHPO	Q43049 schizosacch	372	38	46.3	172	2	Q4UFJ1_THEAN	Q4ufj1 theileria a
300	39	47.6	477	2	Q4WAF0_ASPFU	Q4waf0 aspergillus	373	38	46.3	175	2	Q4V6A8_DROME	Q4v6a8 drosophila
301	39	47.6	480	2	Q7SC42_NEUCR	Q7sc42 neurospora	374	38	46.3	184	1	Q7MBP2_VIBVY	Q7mbp2 vibrio vuln
302	39	47.6	481	2	Q64814_ARATH	Q64814 arabidopsis	375	38	46.3	184	1	YCF4_P5INU	Q8wi09 pilotum nu
303	39	47.6	491	2	Q8VZY3_MUSAC	Q8vzy3 musa acumin	376	38	46.3	191	2	Q8YLC7_RALSO	Q8ylc7 ralstonia s
304	39	47.6	504	2	Q8BLD0_DREHA	Q8blld0 debaryomyce	377	38	46.3	192	2	Q2QXR5_ORYSA	Q2qxr5 oryza sativ
305	39	47.6	505	2	Q2U9U7_ASPOR	Q2u9u7 aspergillus	378	38	46.3	193	2	Q3JCE8_NITOC	Q3jceb nitrosococ
306	39	47.6	507	2	Q8VZY2_MUSAC	Q8vzy2 musa acumin	379	38	46.3	196	2	Q2JR80_9CVAN	Q2jr80 cyanobacter
307	39	47.6	543	2	Q5AU61_EMENI	Q5au61 aspergillus	380	38	46.3	208	1	KTHY_XYLFA	Q9pfef xyella fas
308	39	47.6	545	2	Q07500_YEAST	Q07500 saccharomyc	381	38	46.3	208	1	KTHY_XYLFT	Q87b89 xyella fas
309	39	47.6	547	2	Q3LR03_ARATH	Q3lr03 arabidopsis	382	38	46.3	210	2	Q3H7F3_TRIER	Q3h7f3 trichodesmi
310	39	47.6	554	2	Q2VQI1_RHIME	Q2vqi1 rhizobium m	383	38	46.3	212	2	Q3ZAP6_DROME	Q3zap6 drosophila
311	39	47.6	559	2	Q2U0P2_ASPOR	Q2u0p2 aspergillus	384	38	46.3	214	2	Q7ANV3_BACC1	Q7anv3 bacillus ce
312	39	47.6	564	2	Q4D9U5_TRYCR	Q4d9u5 trypanosoma	385	38	46.3	217	2	Q3R1C3_XYLFA	Q3ric3 xyella fas
313	39	47.6	565	2	Q4IGJ2_GIBZE	Q4igj2 gibberella	386	38	46.3	217	2	Q3R1O1_XYLFA	Q3rio1 xyella fas
314	39	47.6	574	2	Q7S2M5_NEUCR	Q7s2m5 neurospora	387	38	46.3	228	2	Q8UAW6_AGR5	Q8uaw6 agrobacteri
315	39	47.6	577	2	Q53JH2_ORYSA	Q53jh2 oryza sativ	388	38	46.3	229	2	Q5JQ60_HUMAN	Q5jq60 homo sapien
316	39	47.6	593	2	Q6FYV5_BARQU	Q6fyv5 bartonella	389	38	46.3	230	1	PLF1_YEAST	Q4004 saccharomyc
317	39	47.6	603	2	Q585X8_9TRYP	Q585x8 trypanosoma	390	38	46.3	237	2	Q4Y4D8_PLACH	Q4y4d8 plasmodium
318	39	47.6	609	2	Q2SEG6_9GAMM	Q2seg6 habella: che	391	38	46.3	237	2	Q4YE17_PLABE	Q4ye17 plasmodium
319	39	47.6	628	2	Q8YMW6_ANASP	Q8ymv6 anabaena sp	392	38	46.3	238	2	Q2KD43_RH1ET	Q2kd43 rhizobium e
320	39	47.6	637	2	Q3MBC5_ANAVT	Q3mbc5 anabaena va	393	38	46.3	247	1	FABG_MYCBO	P0a5y5 mycobacteri
321	39	47.6	651	2	Q8UBL7_AGR5	Q8ubl7 agrobacteri	394	38	46.3	247	1	FABG_MYCTU	P0a5y4 mycobacteri
322	39	47.6	660	1	SCNNH_XENLA	P51171 xenopus lae	395	38	46.3	247	2	Q3RLB1_RALME	Q3rlb1 ralstonia m
323	39	47.6	661	2	Q353K6_9GAMM	Q353k6 alkalilimni	396	38	46.3	248	2	Q3Q0Z0_9GAMM	Q3q0z0 shewanella

397	38	46.3	248	2	Q700Q0_PSEPU	Q700Q0_pseudomonas	470	38	46.3	578	2	Q8XST7_RALSO	Q8XST7_ralstonia s
398	38	46.3	254	2	Q30RT8_THIDN	Q30rt8 thiomicrosp	471	38	46.3	581	2	Q9LPQ8_ARATH	Q9lpq8 arabidopsis
399	38	46.3	261	2	Q46WJ6_RALEJ	Q46wj6 ralstonia e	472	38	46.3	590	2	Q3WB58_9ACTO	Q3wb58 frankia sp.
400	38	46.3	275	2	Q8UUL7_ORYLA	Q8uul7 oryzias lat	473	38	46.3	600	2	Q8YJX0_ANASP	Q8yjx0 anabaena sp
401	38	46.3	277	2	Q13951_SCHPO	Q13951 schizosacch	474	38	46.3	602	2	Q55MF6_CRYNE	Q55mf6 cryptococcus
402	38	46.3	279	2	Q72VT9_LEPIC	Q72vt9 leptospira	475	38	46.3	607	1	UVRC_PSEPK	Q88fj7 pseudomonas
403	38	46.3	279	2	Q8F9F4_LEPIN	Q8f9f4 leptospira	476	38	46.3	607	2	Q2XCV4_PSEPU	Q2xcv4 pseudomonas
404	38	46.3	279	2	Q8XSG3_RALSO	Q8xsg3 ralstonia s	477	38	46.3	608	2	Q3XV48_9PROT	Q3xv48 magnetococc
405	38	46.3	303	2	Q39R61_GEOMG	Q39r61 geobacter m	478	38	46.3	611	2	Q4DRA4_TRYCR	Q4dra4 trypanosoma
406	38	46.3	305	2	Q73NG5_TREDE	Q73ng5 treponema d	479	38	46.3	611	2	Q4E603_TRYCR	Q4e603 trypanosoma
407	38	46.3	322	2	Q45729_CAEEL	Q45729 caenorhabdi	480	38	46.3	651	2	Q40P39_DESAC	Q40p39 desulfuromo
408	38	46.3	328	2	Q22K23_CALSA	Q22k23 caldicellul	481	38	46.3	652	2	Q8FTM0_CANGA	Q8ftm0 candida gla
409	38	46.3	335	2	Q8ITJ4_HETGL	Q8itj4 heterodera	482	38	46.3	661	2	P73025_SYNY3	P73025 synchocyst
410	38	46.3	343	1	Q7K2T0_DROME	Q7k2t0 drosophila	483	38	46.3	726	1	AP1B1_YEAST	P36000 saccharomyc
411	38	46.3	343	1	ALF_CORGL	P19537 corynebacte	484	38	46.3	746	2	Q2X2S6_9GAMM	Q2x2s6 shewanella
412	38	46.3	344	2	Q2XUR2_CORGL	Q2xur2 corynebacte	485	38	46.3	746	2	Q22WPI_SHEPU	Q22wp1 shewanella
413	38	46.3	344	2	Q6NF15_CORDI	Q6nf15 corynebacte	486	38	46.3	830	2	Q84OC0_9GAMM	Q84oc0 cellivibrio
414	38	46.3	347	2	Q3K6C9_PSEPF	Q3k6c9 pseudomonas	487	38	46.3	843	2	Q8LPN1_ARATH	Q8lpn1 arabidopsis
415	38	46.3	352	2	Q518N7_ENTHI	Q518n7 entamoeba h	488	38	46.3	846	2	Q2ZY28_STRSU	Q2zy28 streptococc
416	38	46.3	356	2	Q9LW71_ARATH	Q9lw71 arabidopsis	489	38	46.3	855	2	Q6BC60_9HIVI	Q6bc60 human immun
417	38	46.3	361	2	Q3FPY7_BURK	Q3fpy7 rhodofera	490	38	46.3	858	2	Q5CKW1_CRYHO	Q5ckw1 cryptospori
418	38	46.3	364	2	Q9W1L8_DROME	Q9w1l8 drosophila	491	38	46.3	864	2	O8ZTU7_PYRAE	Q8ztu7 pyrobaculum
419	38	46.3	367	2	Q93AQ6_BIFLO	Q93aq6 bifidobacte	492	38	46.3	879	2	Q7S4B4_NEUCR	Q7s4b4 neurospora
420	38	46.3	367	2	Q8G6C9_BIFLO	Q8g6c9 bifidobacte	493	38	46.3	883	2	Q4DII4_TRYCR	Q4di44 trypanosoma
421	38	46.3	368	2	Q700D0_ARATH	Q700d0 arabidopsis	494	38	46.3	894	2	Q64X97_BACFR	Q64x97 bacteroides
422	38	46.3	371	2	Q69AA9_ACTPL	Q69aa9 actinobacil	495	38	46.3	935	2	Q4Z0C9_PLABE	Q4z0c9 plasmodium
423	38	46.3	379	2	O81862_ARATH	O81862 arabidopsis	496	38	46.3	942	2	Q89U55_BRAJA	Q89u55 bradyrhizob
424	38	46.3	379	2	Q9ZEX1_BORPE	Q9zex1 bordetella	497	38	46.3	1012	2	Q8S693_ORYSA	Q8s693 oryza aativ
425	38	46.3	379	2	Q7VXJ5_BORPE	Q7vxj5 bordetella	498	38	46.3	1062	2	Q4WY69_ASPFU	Q4wy69 aspergillus
426	38	46.3	379	2	Q7W8Y3_BORPE	Q7w8y3 bordetella	499	38	46.3	1130	2	Q85LJ7_BACLD	Q85lj7 bacillus l1
427	38	46.3	379	2	Q7WKC6_BORBR	Q7wkc6 bordetella	500	38	46.3	1139	2	Q8Q6U7_9GAMR	Q8q6u7 porcine end
428	38	46.3	385	2	Q368D9_9GAMM	Q368d9 shewanella	501	38	46.3	1140	1	Y376_MYCPN	P75405 mycoplasma
429	38	46.3	395	2	O8L5J1_LYCES	O8l5j1 lycopersico	502	38	46.3	1178	2	Q3GDJ2_9FIRM	Q3gdj2 syntrophomo
430	38	46.3	395	2	O8RVL3_LYCES	O8rvl3 lycopersico	503	38	46.3	1183	2	Q3V0E1_MOUSE	Q3v0e1 mus musculu
431	38	46.3	398	2	O81861_ARATH	O81861 arabidopsis	504	38	46.3	1280	2	O84KD1_HORVU	O84kd1 hordeum vul
432	38	46.3	399	2	Q93WT4_LYCES	Q93wt4 lycopersico	505	38	46.3	1290	2	Q4WYPO_ASPFU	Q4wyp0 aspergillus
433	38	46.3	405	2	Q5YFH2_SINGAP	Q5yfh2 singapore g	506	38	46.3	1333	2	Q54W10_DICDI	Q54w10 dictyosteli
434	38	46.3	409	2	Q98G71_RHILO	Q98g71 rhizobium l	507	38	46.3	1526	2	Q9FF08_ARATH	Q9ff08 arabidopsis
435	38	46.3	411	2	Q9FZ29_ARATH	Q9fz29 arabidopsis	508	38	46.3	1540	2	Q9SD90_ARATH	Q9sd90 arabidopsis
436	38	46.3	411	2	Q2JDP9_9ACTO	Q2jdp9 frankia sp.	509	38	46.3	1619	2	Q5CRJ9_CRYPV	Q5crj9 cryptospori
437	38	46.3	411	2	Q820T4_ENTFA	Q820t4 enterococcu	510	38	46.3	1946	2	Q4Y0C5_PLACH	Q4y0c5 plasmodium
438	38	46.3	415	2	O8YVR5_ANASP	O8yvr5 anabaena sp	511	38	46.3	2129	2	Q5QF53_9CAUD	Q5qf53 pseudomonas
439	38	46.3	417	2	Q7Q1L5_ANOGA	Q7q1l5 anophales g	512	38	46.3	2563	2	Q2U3D3_ASPOR	Q2u3d3 aspergillus
440	38	46.3	420	2	Q3M6N9_ANAVT	Q3m6n9 anabaena va	513	38	46.3	2628	2	Q06179_YEAST	Q06179 saccharomyc
441	38	46.3	424	2	Q9EYQ3_CLOCE	Q9eyq3 clostridium	514	38	46.3	3351	1	APLP_DROME	Q9v496 drosophila
442	38	46.3	429	2	Q5Z2C5_NOCFA	Q5z2c5 nocardia fa	515	38	46.3	4020	2	O6ZXI3_MAGGR	Q6zx13 magnaporthe
443	38	46.3	437	2	Q4ZRB4_PSEU2	Q4zrb4 pseudomonas	516	38	46.3	7354	2	Q4PLLS_MOUSE	Q4pll5 mus musculu
444	38	46.3	439	2	O8SAY2_ORYSA	O8say2 oryza sativ	517	37.5	45.7	75	2	Q26728_9TRYP	Q26728 trypanosoma
445	38	46.3	445	2	Q4N176_THEPA	Q4n176 theileria p	518	37.5	45.7	103	2	Q7WXF4_RALEPU	Q7wxf4 ralstonia e
446	38	46.3	458	2	Q3SJR2_THIDA	Q3sjr2 thiobacillu	519	37.5	45.7	176	2	Q9N2G6_PIG	Q9n2g6 sus scrofa
447	38	46.3	462	2	O05090_NOCSE	O05090 nocardioide	520	37.5	45.7	249	2	Q4JA68_SULAC	Q4ja68 sulfolobus
448	38	46.3	464	2	Q53886_SPICI	Q53886 spiroplasma	521	37.5	45.7	267	2	Q5CS98_CRYPV	Q5cs98 cryptospori
449	38	46.3	465	2	Q2SX78_BURTH	Q2sx78 burkholderi	522	37.5	45.7	289	2	Q8IIP9_PLAF7	Q8iip9 plasmodium
450	38	46.3	467	2	O8DLV2_SYNEL	O8dlv2 synechococc	523	37.5	45.7	340	2	Q7IIP9_PLAF7	Q7iip9 cryptospori
451	38	46.3	470	2	Q2W5U2_MAGSA	Q2w5u2 magnetospi	524	37.5	45.7	501	2	Q7M3W4_9TRYP	Q7m3w4 trypanosoma
452	38	46.3	470	2	O6XV41_9NFA	O6xv41 influenza a	525	37.5	45.7	529	2	Q2JX14_9CVAN	Q2jx14 cyanobacter
453	38	46.3	471	2	Q6FW03_CANGA	Q6fw03 candida gla	526	37.5	45.7	542	2	Q2JUT0_9CVAN	Q2jtu0 cyanobacter
454	38	46.3	471	2	O6XV23_TAGU2	Q6xv23 influenza a	527	37.5	45.7	543	2	Q5KC75_CRYNE	Q5kc75 cryptococcu
455	38	46.3	477	2	Q3CJ01_THEET	Q3cj01 thermoanaer	528	37.5	45.7	565	2	Q55J16_CRYNE	Q55j16 cryptococcu
456	38	46.3	481	2	Q3POA7_9GAMM	Q3poa7 shewanella	529	37.5	45.7	701	2	Q9KW03_PSEAE	Q9kw03 pseudomonas
457	38	46.3	481	2	Q3Q9B8_9GAMM	Q3q9b8 shewanella	530	37.5	45.7	775	2	Q91VW8_MOUSE	Q91vw8 mus musculu
458	38	46.3	491	2	Q2ZC63_9GAMM	Q2zc63 shewanella	531	37.5	45.7	779	2	O80ZX2_MOUSE	O80zx2 mus musculu
459	38	46.3	491	2	Q35U27_9GAMM	Q35u27 shewanella	532	37.5	45.7	833	2	Q5CEN0_CRYHO	Q5cen0 cryptospori
460	38	46.3	491	2	Q8EHY2_SHEON	Q8ehy2 shewanella	533	37.5	45.7	926	2	Q5CQR7_CRYPV	Q5cqr7 cryptospori
461	38	46.3	492	2	Q3VT67_PROAE	Q3vt67 prosthecoc	534	37.5	45.7	1038	2	Q4U9D2_THEAN	Q4u9d2 theileria a
462	38	46.3	511	2	Q47GH8_DECAR	Q47gh8 dechloromon	535	37.5	45.7	1134	2	Q50YK0_ENTHI	Q50yk0 entamoeba h
463	38	46.3	517	2	Q2W6L3_MAGSA	Q2w6l3 magnetospi	536	37.5	45.7	1342	2	Q5L1U7_BACFN	Q5l1u7 bacteroides
464	38	46.3	526	2	Q6DCK1_YENLA	Q6dck1 xenopus lae	537	37.5	45.7	1342	2	O64N18_BACFR	O64n18 bacteroides
465	38	46.3	538	2	Q4V902_BRARE	Q4v902 brachydanio	538	37	45.1	46	2	O2KU95_BORAV	O2ku95 bordetella
466	38	46.3	554	2	Q7RKL7_PLAYO	Q7rkl7 plasmodium	539	37	45.1	109	2	O742W6_MYCPA	O742w6 mycobacteri
467	38	46.3	559	2	O9GPT1_DICTDI	O9gpt1 dictyosteli	540	37	45.1	122	2	Q9KSR9_VIBCH	Q9ksr9 vibrio chol
468	38	46.3	561	2	O6CP34_KLUJA	O6cp34 kluyveromyc	541	37	45.1	128	2	P73147_SYNY3	P73147 synchocyst
469	38	46.3	568	2	O6CJY6_KLUJA	O6cjy6 kluyveromyc	542	37	45.1	130	2	Q7SBM0_NEUCR	Q7sbm0 neurospora

543	37	45.1	139	2	Q82RN5	STRAW	Q82rn5	streptomyce	616	37	45.1	362	2	Q731L8	BACC1	Q731l8	bacillus ce
544	37	45.1	140	2	Q4N0Z3	THEPA	Q4n0z3	theileria p	617	37	45.1	362	2	Q81ME9	BACAN	Q81me9	bacillus an
545	37	45.1	143	2	Q3HRY8	SOLITU	Q3hry8	solanum tub	618	37	45.1	363	2	Q9UXJ6	NEIMB	Q9uxj6	neisseria m
546	37	45.1	143	2	Q92X81	RHIME	Q92x81	rhizobium m	619	37	45.1	364	2	Q83HS4	TROM8	Q83hs4	tropheryma
547	37	45.1	146	1	R2Z8	TRYCR			620	37	45.1	365	2	Q47AW0	DECAR	Q47aw0	dechloromon
548	37	45.1	146	2	Q981C3	SULSO	Q981c3	trypanosoma	621	37	45.1	365	2	Q6QW80	AZOB8	Q6qw80	azospirillum
549	37	45.1	146	2	Q4E0P2	TRYCR	Q4e0p2	trypanosoma	622	37	45.1	366	1	HIS8	HALSA	Q6qhs0	halobacteri
550	37	45.1	146	2	Q4E3Q3	TRYCR	Q4e3q3	trypanosoma	623	37	45.1	368	2	Q9AK08	STROO	Q9ak08	streptomyce
551	37	45.1	146	2	Q4E3Q5	TRYCR	Q4e3q5	trypanosoma	624	37	45.1	370	2	Q83GF6	TROWT	Q83gf6	tropheryma
552	37	45.1	150	2	Q7SBB3	NEUCR	Q7sbb3	neurospora	625	37	45.1	371	2	Q5MIY7	ABDAL	Q5miy7	aedes albop
553	37	45.1	156	2	Q82ZJ7	ENTFA	Q82zj7	enterococcu	626	37	45.1	375	2	Q91Z57	MOUSE	Q91z57	mus musculus
554	37	45.1	156	2	Q66091	9ALPH	Q66091	canid herpe	627	37	45.1	376	2	Q5XIK1	RAT	Q5xik1	rattus norv
555	37	45.1	178	2	Q50M19	ENTHI	Q50m19	entamoeba h	628	37	45.1	376	2	Q9D9J3	MOUSE	Q9d9j3	mus musculus
556	37	45.1	190	2	Q6AMA5	DESPS	Q6ama5	desulfotale	629	37	45.1	377	2	Q9F6S8	NEIGO	Q9f6s8	neisseria g
557	37	45.1	194	2	Q4C8S9	CROWT	Q4c8s9	crocosphaer	630	37	45.1	377	2	Q5F7G9	NEIG1	Q5f7g9	neisseria g
558	37	45.1	195	2	Q5QVB4	IDIOE	Q5qvb4	idiomarina	631	37	45.1	379	2	Q6LFP1	PLAF7	Q6lfp1	plasmodium
559	37	45.1	195	2	Q92MN6	RHIME	Q92mn6	rhizobium m	632	37	45.1	382	2	Q3E8P7	ARATH	Q3e8p7	arabidopsis
560	37	45.1	197	2	Q31E08	THICR	Q31e08	thiomicrocep	633	37	45.1	388	2	Q6BY07	DEBHA	Q6by07	debatyomyce
561	37	45.1	197	2	Q7UK86	RHOBA	Q7uk86	rhodopirell	634	37	45.1	389	2	Q53E76	RAT	Q53e76	rattus norv
562	37	45.1	202	2	Q8WPT0	9TRYP	Q8wpt0	trypanosoma	635	37	45.1	395	2	Q43DV4	9CHLB	Q43dv4	chlorobium
563	37	45.1	204	2	Q2NU89	SODGL	Q2nu89	sodalis glo	636	37	45.1	398	2	Q7WE73	BORBR	Q7we73	borderella
564	37	45.1	205	2	Q26683	9TRYP	Q26683	trypanosoma	637	37	45.1	398	2	Q571Q6	SPAAU	Q571q6	sparus aura
565	37	45.1	208	2	Q34FD6	RHOPA	Q34fd6	rhodopseudo	638	37	45.1	401	2	Q303Z4	ARATH	Q303z4	arabidopsis
566	37	45.1	208	2	Q62572	MUSSI	Q62572	mus spicile	639	37	45.1	406	2	Q5SKZ3	THET8	Q5skz3	thermus the
567	37	45.1	219	2	Q96YM7	SULTO	Q96ym7	sulfolobus	640	37	45.1	406	2	Q72LD5	THET2	Q72ld5	thermus the
568	37	45.1	222	2	Q2TZA5	ASFOR	Q2tza5	aspergillus	641	37	45.1	408	2	Q5DL11	9INFA	Q5dl11	influenza a
569	37	45.1	223	2	Q7NJM1	GLOVI	Q7njm1	gloeobacter	642	37	45.1	411	2	Q3W528	9ACTO	Q3w528	frankia sp.
570	37	45.1	227	2	Q62570	MUSSI	Q62570	mus spicile	643	37	45.1	411	2	Q41Y13	DESHA	Q41y13	desulfitoba
571	37	45.1	231	2	Q73IR7	RHOPA	Q73ir7	rhodopseudo	644	37	45.1	414	2	Q6YM50	LYCES	Q6ym50	lycopersico
572	37	45.1	232	1	SNY	MOUSE	P13675	mus musculus	645	37	45.1	415	2	Q9FZ03	LYCES	Q9fz03	lycopersico
573	37	45.1	232	2	Q3TDB	MOUSE	Q3ttd8	m adult mal	646	37	45.1	416	2	Q58P25	CERAE	Q58p25	cercopithec
574	37	45.1	233	2	Q4VMW7	9ROSI	Q4vmw7	spondianthu	647	37	45.1	419	2	Q8CJT2	STROO	Q8cj2	streptomyce
575	37	45.1	249	2	Q4Z5A6	PLABE	Q4zsa6	plasmodium	648	37	45.1	420	1	GSHR	RAT	Q8d2x8	wiggleswort
576	37	45.1	252	2	Q4HQ84	CAMUP	Q4hq84	campylobact	649	37	45.1	424	1	GSHR	RAT	Q70619	rattus norv
577	37	45.1	259	2	Q43MW9	SOLUS	Q43mw9	solibacter	650	37	45.1	425	2	Q503Y5	BRARE	Q503y5	brachydanio
578	37	45.1	259	2	Q80US3	MOUSE	Q80us3	mus musculus	651	37	45.1	426	2	Q43GG3	9CHLB	Q43gg3	chlorobium
579	37	45.1	264	2	Q3KHV4	PSEPF	Q3khv4	pseudomonas	652	37	45.1	429	2	Q88XL3	LACPL	Q88xl3	lactobacill
580	37	45.1	266	2	Q49712	ARATH	Q49712	arabidopsis	653	37	45.1	430	1	DIN7	YEAST	Q82086	clostridum
581	37	45.1	267	2	Q7VCL1	PROMA	Q7vcl1	prochloroco	654	37	45.1	430	2	Q891C3	CLOTE	Q891c3	clostridium
582	37	45.1	269	2	Q33S85	9GAMT	Q33s85	shewanella	655	37	45.1	431	2	Q4K8X8	PSEF5	Q4k8x8	pseudomonas
583	37	45.1	271	2	Q3CMN5	ALTAT	Q3cmn5	pseudocalter	656	37	45.1	433	2	Q98NF5	RHIL0	Q98nf5	rhizobium l
584	37	45.1	271	2	Q4AF60	9CHLB	Q4af60	chlorobium	657	37	45.1	435	2	Q31YC7	RHOS4	Q31yc7	rhodobacter
585	37	45.1	275	1	ATG3	SCHPO	Q43035	schizosacch	658	37	45.1	436	2	Q7U6B2	SYNPX	Q7u6b2	synechococc
586	37	45.1	277	2	Q8GCX3	ENTMU	Q8gcx3	enterococcu	659	37	45.1	437	2	Q48GU6	PSEL4	Q48gu6	pseudomonas
587	37	45.1	278	2	Q57EL4	BRUAB	Q57el4	brucella ab	660	37	45.1	437	2	Q7D648	MYCTU	Q7d648	mycobacteri
588	37	45.1	278	2	Q8G227	BRUSU	Q8g227	brucella su	661	37	45.1	437	2	P77909	MYCTU	P77909	mycobacteri
589	37	45.1	278	2	Q8YFU2	BRUME	Q8yfu2	brucella me	662	37	45.1	437	2	Q7TX94	MYCBO	Q7tx94	mycobacteri
590	37	45.1	278	2	Q2YMM1	BRUA2	Q2ymml	brucella ab	663	37	45.1	437	2	Q87Z96	PSSSM	Q87z96	pseudomonas
591	37	45.1	283	1	CBP3	SCHPO	Q9usK6	schizosacch	664	37	45.1	441	2	Q4HVE2	GIBZE	Q4hve2	gibberella
592	37	45.1	284	2	Q6W1X4	RHISN	Q6w1x4	rhizobium s	665	37	45.1	442	2	Q91IN2	PSEAB	Q91in2	pseudomonas
593	37	45.1	293	2	Q3B2L7	PELLD	Q3b2l7	pelodictyon	666	37	45.1	444	2	Q7Q0F9	ANOQA	Q7q0f9	anopheles g
594	37	45.1	301	2	Q448T3	SOLUS	Q448t3	solibacter	667	37	45.1	445	2	Q46CU3	METBA	Q46cu3	methanosaar
595	37	45.1	305	2	Q3XPX4	9PROT	Q3xpx4	magnetococc	668	37	45.1	445	2	Q61JA2	CABBR	Q61ja2	caenorhabdi
596	37	45.1	306	2	Q67NJ6	SYWTH	Q67nj6	symbiobacte	669	37	45.1	446	2	Q09646	CASEL	Q09646	caenorhabdi
597	37	45.1	306	2	Q74DC4	GROSL	Q74dc4	geobacter s	670	37	45.1	446	2	Q6XV50	9INFA	Q6xv50	influenza a
598	37	45.1	307	2	Q5ZUA8	LEGPH	Q5zuu8	legionella	671	37	45.1	453	2	Q3U9R7	MOUSE	Q3u9r7	mus musculus
599	37	45.1	310	1	VGR1	YEAST	P53102	saccharomyc	672	37	45.1	454	2	Q87HA7	VIBPA	Q87ha7	vibrio para
600	37	45.1	316	2	Q5WVF2	LEGPL	Q5wvf2	legionella	673	37	45.1	457	2	Q4MXE3	BACCE	Q4mx3	bacillus ce
601	37	45.1	316	2	Q5X409	LEGPA	Q5x409	legionella	674	37	45.1	458	2	Q3KEE1	PSEPF	Q3kee1	pseudomonas
602	37	45.1	318	2	Q46E04	METBA	Q46e04	methanosaar	675	37	45.1	458	2	Q4KE85	PSEF5	Q4kee5	pseudomonas
603	37	45.1	324	2	Q6KZ46	PICTO	Q6kz46	picrophilus	676	37	45.1	460	2	Q36YH2	RHOPA	Q36yh2	rhodopseudo
604	37	45.1	327	2	Q2MF05	STRTN	Q2mf05	streptomyce	677	37	45.1	465	2	Q9AU19	XANSI	Q9au19	xanthorhiza
605	37	45.1	330	2	Q3C6K9	9CLOT	Q3c6k9	alkaliphiliu	678	37	45.1	465	2	Q9AU21	9MAGN	Q9au21	tetracentro
606	37	45.1	333	2	Q4NC91	9MICC	Q4nc91	arthrobacte	679	37	45.1	465	2	Q8BW56	MOUSE	Q8bw56	mus musculus
607	37	45.1	338	2	Q6BA55	9PROT	Q6ba55	uncultured	680	37	45.1	470	2	Q2W9Y4	MAGSA	Q2w9y4	magnetospir
608	37	45.1	341	2	Q3C4C1	9CLOT	Q3c4c1	alkaliphiliu	681	37	45.1	470	2	Q6DWT5	9INFA	Q6dwt5	influenza a
609	37	45.1	342	2	Q9YQ33	ANASP	Q9yq33	anabaena sp	682	37	45.1	470	2	Q6DX23	9INFA	Q6dx23	influenza a
610	37	45.1	345	1	ALF	MYCLE	Q69600	mycobacteri	683	37	45.1	470	2	Q6DX31	9INFA	Q6dx31	influenza a
611	37	45.1	348	2	Q4ML61	BACCE	Q4ml61	bacillus ce	684	37	45.1	470	2	Q6DXB9	9INFA	Q6dxb9	influenza a
612	37	45.1	351	2	Q564V4	CABEL	Q564v4	caenorhabdi	685	37	45.1	470	2	Q6DXD5	9INFA	Q6dxd5	influenza a
613	37	45.1	353	2	Q8YL93	ANASP	Q8yl93	anabaena sp	686	37	45.1	470	2	Q6DXF3	9INFA	Q6dxf3	influenza a
614	37	45.1	355	2	Q4IEH9	GIBBEZ	Q4ieh9	gibberella	687	37	45.1	470	2	Q6DXG7	9INFA	Q6dxg7	influenza a
615	37	45.1	362	2	Q3ETB4	BACTI	Q3etb4	bacillus th	688	37	45.1	470	2	Q6XV19	9INFA	Q6xv19	influenza a

689	37	45.1	470	2	Q6XV20_9INFA	Q6xv20 influenza a	762	37	45.1	654	2	Q9HZC6_PSEAE	Q9hzc6 pseudomonas
690	37	45.1	470	2	Q6XV21_9INFA	Q6xv21 influenza a	763	37	45.1	655	2	Q9PR74_UREPA	Q9pr74 ureaplasma
691	37	45.1	470	2	Q6XV22_9INFA	Q6xv22 influenza a	764	37	45.1	683	2	Q8GH80_DICDI	Q8gh80 dictyostelli
692	37	45.1	470	2	Q6XV24_TAGUA	Q6xv24 influenza a	765	37	45.1	710	2	Q3L955_RHOE4	Q3l955 rhodococcus
693	37	45.1	470	2	Q6XV25_9INFA	Q6xv25 influenza a	766	37	45.1	748	2	Q2QPL4_ORYSA	Q2qpl4 oryza sativ
694	37	45.1	470	2	Q6XV26_TADNZ	Q6xv26 influenza a	767	37	45.1	752	2	Q2QPK8_ORYSA	Q2qpk8 oryza sativ
695	37	45.1	470	2	Q6XV27_9INFA	Q6xv27 influenza a	768	37	45.1	754	2	Q9SR95_ARATH	Q9sr95 arabidopsis
696	37	45.1	470	2	Q6XV28_9INFA	Q6xv28 influenza a	769	37	45.1	759	2	Q05143_RUMFL	Q05143 ruminococcu
697	37	45.1	470	2	Q6XV29_9INFA	Q6xv29 influenza a	770	37	45.1	765	2	Q4P0C9_USTMA	Q4p0c9 ustilago ma
698	37	45.1	470	2	Q6XV40_9INFA	Q6xv40 influenza a	771	37	45.1	768	2	Q23638_ARATH	Q23638 arabidopsis
699	37	45.1	470	2	Q9E147_9INFA	Q9e147 influenza a	772	37	45.1	777	2	Q3EC72_ARATH	Q3ec72 arabidopsis
700	37	45.1	472	2	Q9LEQ6_ARATH	Q9leq6 arabidopsis	773	37	45.1	778	2	Q662r4 BORGA	Q662r4 borrella ga
701	37	45.1	473	2	Q4FLI6_PELUB	Q4fli6 pelagibacte	774	37	45.1	780	2	Q8LP50_ARATH	Q8lp50 arabidopsis
702	37	45.1	477	2	Q9XID1_ARATH	Q9xid1 arabidopsis	775	37	45.1	780	2	Q9SIF2_ARATH	Q9sif2 arabidopsis
703	37	45.1	478	2	Q4XW7 ASPFU	Q4xw7 aspergillus	776	37	45.1	781	2	Q43638_SECEC	Q43638 secale cere
704	37	45.1	480	2	Q7VDO8_PROMA	Q7vdg8 prochloroco	777	37	45.1	785	2	Q6ZCV7_ORYSA	Q6zcv7 oryza sativ
705	37	45.1	482	2	Q3UAS7_MOUSE	Q3uas7 mus musculu	778	37	45.1	794	2	Q6YW98_ORYSA	Q6yw98 oryza sativ
706	37	45.1	482	2	Q3UB84_MOUSE	Q3ub84 mus musculu	779	37	45.1	796	2	Q5ID24_BRARE	Q5id24 brachydanio
707	37	45.1	483	2	Q3TXK9_MOUSE	Q3txk9 mus musculu	780	37	45.1	799	2	Q54Q50_DICDI	Q54q50 dictyostelli
708	37	45.1	491	2	Q3TWI5_MOUSE	Q3twi5 mus musculu	781	37	45.1	802	2	Q4UFR7_THEAN	Q4ufr7 theileria a
709	37	45.1	493	2	Q6CGB7_YARLI	Q6cgb7 yarrowia li	782	37	45.1	804	1	SYL_STAAM	P67512 staphylococ
710	37	45.1	493	2	Q442G5_SOLUS	Q442g5 solibacter	783	37	45.1	804	1	SYL_STAAN	P67513 staphylococ
711	37	45.1	500	1	GSRR_MOUSE	P47791 mus musculu	784	37	45.1	804	1	SYL_STAAR	Q6gfu3 staphylococ
712	37	45.1	505	1	GATB_STRAW	Q82j12 streptomyce	785	37	45.1	804	2	Q2YTH9_STAAB	Q2ych9 staphylococ
713	37	45.1	511	2	Q6BYF0_DEBHA	Q6byf0 debaryomyce	786	37	45.1	805	1	SYL_STAAC	Q5hi16 staphylococ
714	37	45.1	513	2	Q7RHW9_PLAYO	Q7rhw9 plasmodium	787	37	45.1	806	1	SYL_STAAS	Q6g8g9 staphylococ
715	37	45.1	514	2	Q5MY91_HORSE	Q5my91 equus cabal	788	37	45.1	807	1	SYL_STAAN	Q8nw17 staphylococ
716	37	45.1	514	2	Q3FDI4_9BURK	Q3fdi4 burkholderi	789	37	45.1	810	2	Q4N3Q3_THEPA	Q4n3q3 theileria p
717	37	45.1	515	2	Q3QUS2_RRHOB	Q3qus2 silicibacte	790	37	45.1	816	2	Q4YAW4_PLABE	Q4yaw4 plasmodium
718	37	45.1	526	2	Q5SNA6_ORYSA	Q5sna6 oryza sativ	791	37	45.1	819	1	PUPA_PSEPU	P25184 pseudomonas
719	37	45.1	529	2	Q92JT2_RHIME	Q92jt2 rhizobium m	792	37	45.1	819	2	Q4GH96_9DEIO	Q4gh96 deinococcus
720	37	45.1	536	1	HPSE3_ARATH	Q9fzp1 arabidopsis	793	37	45.1	834	2	Q6G2A3_BARHE	Q6g2a3 bartonella
721	37	45.1	537	2	Q70YJ3_HORVU	Q70yj3 hordeum vul	794	37	45.1	844	1	HEXA_STRRG	P0a3r3 streptococc
722	37	45.1	538	1	RT03_ORYSA	P46773 oryza sativ	795	37	45.1	844	1	HEXA_STRRG	P0a3r4 streptococc
723	37	45.1	538	1	Q7JAJ0_ORYSA	Q7jaj0 oryza sativ	796	37	45.1	849	2	Q8RDR3_FUSNN	Q8rdr3 fusbacteri
724	37	45.1	539	1	HPSE2_ARATH	Q81608 arabidopsis	797	37	45.1	857	2	Q746A8_THET2	Q746a8 thermus the
725	37	45.1	539	2	Q2QN56_ORYSA	Q2qn56 oryza sativ	798	37	45.1	874	2	Q7EFA0_BRARE	Q7efao brachydanio
726	37	45.1	541	2	Q4WGH9_ASPFU	Q4wgh9 aspergillus	799	37	45.1	892	2	Q6EF99_BRARE	Q6ef99 brachydanio
727	37	45.1	541	2	Q69I16_ORYSA	Q69i16 oryza sativ	800	37	45.1	900	2	Q4OX07_KINRA	Q4ox07 kinococcus
728	37	45.1	541	2	Q98RH0_MYCPU	Q98rh0 mycoplasma	801	37	45.1	905	2	Q3SVV1_NITWN	Q3svv1 nitrobacter
729	37	45.1	542	2	Q7XB94_ORYSA	Q7xb94 oryza sativ	802	37	45.1	909	2	Q2ZII15_CALSA	Q2zii15 caldicellul
730	37	45.1	542	2	Q94HK9_ORYSA	Q94hk9 oryza sativ	803	37	45.1	927	2	Q7XH20_ORYSA	Q7xh20 oryza sativ
731	37	45.1	542	2	Q34W08_GGAMW	Q34wq8 alkalilimni	804	37	45.1	927	2	Q9AYH9_ORYSA	Q9ayh9 oryza sativ
732	37	45.1	543	1	HPSE1_ARATH	Q9if10 arabidopsis	805	37	45.1	953	2	Q5LAN5_BACFN	Q5lan5 bacteroides
733	37	45.1	550	2	Q3SHZ1_THIDA	Q3shz1 thiobacillu	806	37	45.1	953	2	Q64R33_BACFR	Q64r33 bacteroides
734	37	45.1	559	2	Q94I86_TOBAN	Q94i86 nicotiana t	807	37	45.1	960	2	Q5OWG0_ENTHI	Q5owg0 entamoeba h
735	37	45.1	562	2	Q4UH35_THEAN	Q4uh35 theileria a	808	37	45.1	974	2	Q4N7P5_THEPA	Q4n7p5 theileria p
736	37	45.1	563	2	Q5QJ44_9SOLA	Q5qj44 nicotiana a	809	37	45.1	993	2	Q488K9_COLP3	Q488k9 colwellia p
737	37	45.1	564	2	Q9KR28_VIBCH	Q9kr28 vibrio chol	810	37	45.1	1058	2	Q4P8Y2_USTMA	Q4p8y2 ustilago ma
738	37	45.1	568	2	Q2UDP9_ASPOR	Q2ujd9 aspergillus	811	37	45.1	1070	2	Q9EZV7_PASMU	Q9ezv7 pasteurella
739	37	45.1	573	2	Q2U380_ASPOR	Q2u380 aspergillus	812	37	45.1	1075	2	Q9M113_ARATH	Q9m113 arabidopsis
740	37	45.1	576	2	Q4N813_THEPA	Q4n813 theileria p	813	37	45.1	1087	2	Q54L14_DICDI	Q54l14 dictyostelli
741	37	45.1	579	2	Q7UZK5_PROMP	Q7uzk5 prochloroco	814	37	45.1	1099	2	Q6EFA1_BRARE	Q6efai brachydanio
742	37	45.1	579	2	Q6DI23_BRARE	Q6di23 brachydanio	815	37	45.1	1100	2	Q5TTS1_ANOGA	Q5tts1 anopheles g
743	37	45.1	585	2	Q7RLN4_PLAYO	Q7rln4 plasmodium	816	37	45.1	1107	2	Q8RWQ2_ARATH	Q8rwq2 arabidopsis
744	37	45.1	587	1	GOX_PENAG	P81156 penicillium	817	37	45.1	1108	2	Q2V3P5_ARATH	Q2v3p5 arabidopsis
745	37	45.1	587	2	Q748B7_GEOSL	Q748b7 geobacter s	818	37	45.1	1110	2	Q8A894_BACTN	Q8a894 bacteroides
746	37	45.1	605	1	GOX_TALFL	Q92452 talaromyces	819	37	45.1	1138	2	Q7QC66_ANOGA	Q7qc66 anopheles g
747	37	45.1	605	2	Q9URJ8_PENAG	Q9urj8 penicillium	820	37	45.1	1150	2	Q01913_CAEEL	Q01913 caenorhabd
748	37	45.1	611	2	Q5GME7_9TELE	Q5gme7 chirocentru	821	37	45.1	1160	2	Q4SFB1_TETNG	Q4sfb1 tetraodon n
749	37	45.1	618	2	Q3EUM3_BACTI	Q3eum3 bacillus th	822	37	45.1	1187	2	Q4WC15_ASPFU	Q4wc15 aspergillus
750	37	45.1	618	2	Q4MNF9_BACCE	Q4mnf9 bacillus ce	823	37	45.1	1229	2	Q5B798_EMENI	Q5b798 aspergillus
751	37	45.1	618	2	Q6HFJ4_BACHK	Q6hjf4 bacillus th	824	37	45.1	1594	2	Q51DK0_ENTHI	Q51dk0 entamoeba h
752	37	45.1	618	2	Q81AG5_BACCR	Q81ag5 bacillus ce	825	37	45.1	1661	2	Q23557_ARATH	Q23557 arabidopsis
753	37	45.1	618	2	Q81Y93_BACAN	Q81y93 bacillus an	826	37	45.1	1662	2	Q9VV68_DROME	Q9vv68 drosophila
754	37	45.1	619	2	Q733N6_BACCI	Q733n6 bacillus ce	827	37	45.1	1936	2	Q4ECV6_BURVI	Q4ecv6 burkholderi
755	37	45.1	621	2	Q65X23_ORYSA	Q65x23 oryza sativ	828	37	45.1	2426	2	Q96UR0_ASPFU	Q96ur0 aspergillus
756	37	45.1	631	2	Q48043_HAEIN	Q48043 haemophilus	829	37	45.1	2426	2	Q4X143_ASPFU	Q4x143 aspergillus
757	37	45.1	639	2	Q9W1J9_DROME	Q9w1j9 drosophila	830	37	45.1	2453	2	Q7R1N7_GIALA	Q7r1n7 giardia lam
758	37	45.1	652	2	Q47A15_DECAR	Q47a15 dechloromon	831	37	45.1	4251	2	Q4RFS9_TETNG	Q4rf54 tetraodon n
759	37	45.1	654	2	Q6UWU2_HUMAN	Q6uwu2 homo sapien	832	36.5	44.5	170	2	Q9QJK9_5VIRU	Q9qjk9 banana bunc
760	37	45.1	654	2	Q95LV1_WACFA	Q95lv1 macaca fasc	833	36.5	44.5	171	2	Q45T42_CLODI	Q45t42 clostridium
761	37	45.1	654	2	Q95LV1_WACFA	Q95lv1 macaca fasc	834	36.5	44.5	182	2	Q3CEN7_THEET	Q3cen7 thermoaer



835	36.5	44.5	325	2	Q83T14	CLOSTRIDIUM	Q83T14	clostridium	908	36	43.9	215	2	Q86G71	DERVA	Q86G71	dermaceator	908
836	36.5	44.5	344	2	Q4MZN4	THEPRA	Q4mzn4	theileria p	909	36	43.9	220	2	Q7PGQ5	ANOGEA	Q7pgq5	anopheles g	909
837	36.5	44.5	348	2	Q3GDT7	9ETIRM	Q3gdt7	syntrophomo	910	36	43.9	222	1	DSBA	ENTAM	Q9xdp1	enterobacte	910
838	36.5	44.5	450	1	GLMM_AGR75		Q8u919	agrobacteri	911	36	43.9	225	1	FLIH	BUCPB	Q89a28	buchnera ap	911
839	36.5	44.5	450	2	Q2K4M3	RH1ET	Q2k4m3	rhizobium e	912	36	43.9	227	2	Q4CO19	CROWT	Q4co19	crocophae	912
840	36.5	44.5	451	1	GLMM_BRUSE		Q8yiu8	brucella me	913	36	43.9	229	2	Q6PL57	ELECO	Q6pl57	eleusine co	913
841	36.5	44.5	451	1	GLMM_BRUSE		Q8fz13	brucella su	914	36	43.9	229	2	Q8X5C5	ECOS7	Q8x5c5	escherichia	914
842	36.5	44.5	451	2	Q57BJ0	BRUAB	Q57bj0	brucella ab	915	36	43.9	233	2	Q88UP6	LACPL	Q88up6	lactobacill	915
843	36.5	44.5	451	2	Q2YQHR	BRUA2	Q2yqhr	brucella ab	916	36	43.9	235	2	Q96174	PLAF7	Q96174	plasmodium	916
844	36.5	44.5	460	2	Q5CKB0	CRYHO	Q5ckb0	cryptospori	917	36	43.9	235	2	Q2WN99	CLOBE	Q2wn99	clostridium	917
845	36.5	44.5	461	2	Q5CT31	CRYPV	Q5ct31	cryptospori	918	36	43.9	245	2	Q2NIZ3	9MOLU	Q2niz3	aster yello	918
846	36.5	44.5	465	2	Q8DTA8	STRMU	Q8dta8	streptococc	919	36	43.9	245	2	Q6QYV5	ONYPE	Q6qyv5	onion yello	919
847	36.5	44.5	475	2	Q3YOV8	ENTFC	Q3yov8	enterococc	920	36	43.9	251	2	Q44JT8	CHRSL	Q44jt8	chromohalob	920
848	36.5	44.5	506	2	Q3SDX4	PARTF	Q3sdx4	paramecium	921	36	43.9	253	2	Q7WTF6	9ACTO	Q7wtf6	streptomyce	921
849	36.5	44.5	509	2	Q3SCX6	PARTF	Q3scx6	paramecium	922	36	43.9	254	2	Q742F0	MYCPA	Q742f0	mycobacteri	922
850	36.5	44.5	560	2	Q6AB58	PROAC	Q6ab58	propionibac	923	36	43.9	259	2	Q61YY3	CABER	Q61yy3	caenorhabdi	923
851	36.5	44.5	579	2	Q3N2K0	9DELT	Q3n2k0	syntrophoba	924	36	43.9	261	1	Y202	ENCCU	Q8et35	encephalito	924
852	36.5	44.5	643	2	Q5VMB3	ORYSA	Q5vmb3	oryza sativ	925	36	43.9	261	1	Q361Z1	MARHY	Q361z1	marinobacte	925
853	36.5	44.5	648	2	Q6NC95	RHOPA	Q6nc95	rhodopseudo	926	36	43.9	263	1	KCTD2	HUMAN	Q14681	homo sapien	926
854	36.5	44.5	720	2	Q37DB6	RHOPA	Q37db6	rhodopseudo	927	36	43.9	263	2	Q51716	BACNO	Q51716	bacteroides	927
855	36.5	44.5	728	2	Q6XQB2	9CAUD	Q6xqb2	enterobacte	928	36	43.9	263	2	Q4VBE7	MOUSE	Q4vbe7	mus muscucu	928
856	36.5	44.5	734	2	Q35DA6	9BRAD	Q35da6	bradyrhizob	929	36	43.9	264	2	Q5KRH7	CORGL	Q5krh7	corynebacte	929
857	36.5	44.5	734	2	Q379S9	RHOPA	Q379s9	rhodopseudo	930	36	43.9	265	2	Q8NZZ4	STRP8	Q8nzz4	streptococc	930
858	36.5	44.5	899	2	Q3ILA4	PSEHT	Q3ila4	pseudocalter	931	36	43.9	266	1	KCTD2	MOUSE	Q8ce20	mus muscucu	931
859	36.5	44.5	928	2	Q4CCP8	CLOTM	Q4ccp8	c clostridi	932	36	43.9	271	2	Q3C8G2	9CLOT	Q3c8g2	alkaliphilu	932
860	36.5	44.5	954	2	Q4B5Q4	9BURK	Q4b5q4	polaromonas	933	36	43.9	272	2	Q9HAE5	HUMAN	Q9hae5	homo sapien	933
861	36	43.9	35	2	Q97RG6	STRPN	Q97rg6	streptococc	934	36	43.9	272	2	Q9LYV6	ARATH	Q9lyv6	arabidopais	934
862	36	43.9	73	2	Q5WFV9	BACSK	Q5wfv9	bacillus cl	935	36	43.9	273	2	Q3CYV8	STRAG	Q3cyv8	streptococc	935
863	36	43.9	79	2	Q5WD50	BACSK	Q5wd50	bacillus cl	936	36	43.9	273	2	Q3D5M2	STRAG	Q3d5m2	streptococc	936
864	36	43.9	91	2	Q8GTA2	DATFE	Q8gta2	datura fero	937	36	43.9	273	2	Q3DIW0	STRAG	Q3diw0	streptococc	937
865	36	43.9	93	2	Q4XAN3	PLACH	Q4xan3	plasmodium	938	36	43.9	273	2	Q3K2X4	STRAL	Q3k2x4	streptococc	938
866	36	43.9	112	2	Q9P117	HUMAN	Q9p117	homo sapien	939	36	43.9	273	2	Q8E1E6	STRAS	Q8e1e6	streptococc	939
867	36	43.9	115	2	Q49UG6	STASJ	Q49ug6	staphylococ	940	36	43.9	273	2	Q8B6W3	STRAG	Q8b6w3	streptococc	940
868	36	43.9	115	2	Q4LAA4	STAHJ	Q4laa4	staphylococ	941	36	43.9	277	2	Q6RVX9	BIFAN	Q6rvx9	bifidobacte	941
869	36	43.9	116	2	Q5A0A4	CANAL	Q5a0a4	candida alb	942	36	43.9	278	2	Q2Y6N4	NITMU	Q2y6n4	nitrosospir	942
870	36	43.9	120	2	Q41XR8	DESHA	Q41xr8	deaultitoba	943	36	43.9	281	2	Q9HPJ3	HALSA	Q9hpj3	halobacteri	943
871	36	43.9	125	2	Q9D3W9	MOUSE	Q9d3w9	mus muscucu	944	36	43.9	281	2	Q6LKN1	PHOPR	Q6lkn1	photobacter	944
872	36	43.9	130	2	Q2SHU9	9GAMM	Q2shu9	hahella che	945	36	43.9	283	2	Q5ICQ7	ENTHI	Q5icq7	entamoeba h	945
873	36	43.9	131	2	Q8TFH6	SCHPO	Q8tfh6	schizosacch	946	36	43.9	284	2	Q7V0Q4	PROMP	Q7v0q4	prochloroco	946
874	36	43.9	131	2	Q6IL14	DROME	Q6il14	drosophila	947	36	43.9	286	2	Q6MRJ0	BDEBA	Q6mrj0	bdellovibri	947
875	36	43.9	138	2	P73331	SYNY3	P73331	synechocyst	948	36	43.9	288	1	DJLA	HAEN	Q46070	haemophilus	948
876	36	43.9	138	2	P73781	SYNY3	P73781	synechocyst	949	36	43.9	288	2	Q2WJ64	CLOBE	Q2wj64	clostridium	949
877	36	43.9	138	2	P73924	SYNY3	P73924	synechocyst	950	36	43.9	288	2	Q4QNR8	HAS18	Q4qnr8	haemophilus	950
878	36	43.9	139	2	Q2ISY0	RHOPA	Q2isy0	rhodopseudo	951	36	43.9	291	2	Q5L3S5	GEOKA	Q5l3s5	geobacillus	951
879	36	43.9	140	2	Q2P3J1	XANOR	Q2p3j1	xanthomonas	952	36	43.9	292	2	Q692L5	YEREN	Q692l5	yersinia en	952
880	36	43.9	144	2	Q4XUE7	PLACH	Q4xue7	plasmodium	953	36	43.9	294	2	Q39YH9	GEOMG	Q39yh9	geobacter m	953
881	36	43.9	144	2	Q8VUB4	9BRAD	Q8vub4	bradyrhizob	954	36	43.9	295	2	Q3DCG8	STRAG	Q3dcb8	streptococc	954
882	36	43.9	146	2	Q7G6B0	ORYSA	Q7g6b0	oryza sativ	955	36	43.9	295	2	Q3DQH3	STRAG	Q3dqh3	streptococc	955
883	36	43.9	147	2	Q34BK2	RHOPA	Q34bk2	rhodopseudo	956	36	43.9	301	2	Q4FSES	PSYAR	Q4fses	psychobact	956
884	36	43.9	152	1	G12	BRARE	P32002	brachydanio	957	36	43.9	302	2	Q82Q71	STRAM	Q82q71	streptomyce	957
885	36	43.9	154	1	SRCB	BUCAP	P32002	buchnera ap	958	36	43.9	302	2	Q3NRS4	SHEPR	Q3nrs4	shewanella	958
886	36	43.9	154	2	Q40JE3	HRCH	Q40je3	ehrlichia c	959	36	43.9	303	1	PECR	RAT	Q9wvk3	rattus norv	959
887	36	43.9	155	2	Q5CPL5	CRYPV	Q5cpl5	cryptospori	960	36	43.9	306	2	Q46HJ6	PROMT	Q46hj6	prochloroco	960
888	36	43.9	156	2	Q9YME3	9HIVI	Q9yme3	human immun	961	36	43.9	311	2	Q2IXS5	RHOPA	Q2ixs5	rhodopseudo	961
889	36	43.9	171	2	Q441Y2	SOLUS	Q441y2	solibacter	962	36	43.9	312	2	Q4PFE9	USTMA	Q4pfe9	ustaliago ma	962
890	36	43.9	172	2	Q7MXP8	PORG1	Q7mxp8	porphyromon	963	36	43.9	313	2	Q6BV50	DEBHA	Q6bv50	debfiarymce	963
891	36	43.9	173	1	FLIV	SALMU	P37588	salmonella	964	36	43.9	313	2	Q5CJF2	CRYHO	Q5cjf2	cryptospori	964
892	36	43.9	174	2	Q17BN5	RHOPA	Q37bm5	rhodopseudo	965	36	43.9	313	2	Q5CVC3	CRYPV	Q5cvc3	cryptospori	965
893	36	43.9	175	2	Q2W619	MAGSA	Q2w619	magnetospir	966	36	43.9	314	2	Q9FI45	ARATH	Q9fi45	arabidopais	966
894	36	43.9	175	2	Q3GLB2	9GAMM	Q3glb2	psychrobact	967	36	43.9	315	2	Q4EMN8	LISMO	Q4emn8	listeria mo	967
895	36	43.9	177	2	Q8F1G8	ECOL6	Q8f1g8	escherichia	968	36	43.9	315	2	Q927U3	LISIN	Q927u3	listeria in	968
896	36	43.9	183	2	Q86K28	DICDI	Q86k28	dictyosteli	969	36	43.9	315	2	Q93RN0	LISMO	Q93rn0	listeria mo	969
897	36	43.9	183	2	Q3LNS6	ARATH	Q9lns6	arabidopais	970	36	43.9	316	2	Q4S6L2	TETNG	Q4s6l2	tetradodon n	970
898	36	43.9	194	2	Q3IF16	PSEHT	Q3ifi6	pseudocalter	971	36	43.9	323	2	Q62050	CAEEL	Q62050	caenorhabdi	971
899	36	43.9	201	2	Q4UDH8	THEAN	Q4udh8	theileria a	972	36	43.9	323	2	Q3YS21	EHRCJ	Q3yse21	ehrlichia c	972
900	36	43.9	201	2	Q9ASV1	ARATH	Q9asv1	arabidopais	973	36	43.9	324	2	Q7QH17	ANOGEA	Q7qh17	anopheles g	973
901	36	43.9	201	2	Q3HF65	TRIER	Q3hf65	trichodesmi	974	36	43.9	324	2	Q7VXB3	BORPE	Q7vxb3	bordetella	974
902	36	43.9	203	2	Q97YK0	SULSO	Q97yk0	sulfolobus	975	36	43.9	325	2	Q3KZ21	BIFAN	Q3kz21	bifidobacte	975
903	36	43.9	207	2	Q6FU74	CANGA	Q6fu74	candida gla	976	36	43.9	326	2	Q5CU30	CRYPV	Q5cu30	cryptospori	976
904	36	43.9	209	2	Q7AF16	ECO57	Q7af16	escherichia	977	36	43.9	326	2	Q45802	BACVU	Q45802	bacteroides	977
905	36	43.9	212	2	Q54550	HABIN	Q54550	haemophilus	978	36	43.9	327	2	Q4DVZ0	TRYCR	Q4dvz0	trypanosoma	978
906	36	43.9	214	2	Q3MPV5	CANAL	Q3mpv5	candida alb	979	36	43.9	327	2	Q4N945	THEPRA	Q4n945	theileria p	979
907	36	43.9	214	2	Q5AFH0	CANAL	Q5afh0	candida alb	980	36	43.9	329	2	Q3JT23	BURP1	Q3jt23	burkholderi	980

981 36 43.-9 331 2 Q3EXT7\_BACTI Q3ext7 bacillus th  
 982 36 43.-9 335 2 Q18363\_CAEL Q18363 caenothabi  
 983 36 43.-9 338 1 AMIF\_PSSM Q887d9 pseudomonas  
 984 36 43.-9 338 2 Q42X2\_PSEU2 Q42xa2 pseudomonas  
 985 36 43.-9 342 2 Q4B4K4\_9BURK Q4b4k4 polaromonas  
 986 36 43.-9 343 2 Q3FVW2\_9BURK Q3fvm2 rhodotera  
 987 36 43.-9 344 2 Q2X9S7\_PSEPU Q2x9s7 pseudomonas  
 988 36 43.-9 344 2 Q88Q56\_PSEPU Q88q56 pseudomonas  
 989 36 43.-9 345 2 Q46DR8\_METBA Q46dr8 methanosarc  
 990 36 43.-9 345 2 Q8TUM6\_METAC Q8tum6 methanosarc  
 991 36 43.-9 345 2 Q3F5E9\_9BURK Q3f5e9 burkholderi  
 992 36 43.-9 345 2 Q40H34\_9RHOB Q40h34 jannaschia  
 993 36 43.-9 346 2 Q8XYM5\_RALSO Q8xym5 ralstonia s  
 994 36 43.-9 347 2 Q2X7Q0\_9GAMM Q2x7q0 shewanella  
 995 36 43.-9 347 2 Q3Q5X6\_9GAMM Q3q5x6 shewanella  
 996 36 43.-9 348 2 Q4XLK2\_FLACH Q4xlk2 plasmodium  
 997 36 43.-9 348 2 Q3FG54\_9BURK Q3fgz4 burkholderi  
 998 36 43.-9 348 2 Q44XX9\_9BURK Q44xx9 burkholderi  
 999 36 43.-9 348 2 Q4BA78\_BURVI Q4ba78 burkholderi  
 1000 36 43.-9 348 2 Q4LR32\_9BURK Q4lr32 burkholderi

## ALIGNMENTS

RESULT 1  
 HPSE\_HUMAN STANDARD; PRT; 543 AA.  
 AC Q9Y251; Q53GE5; Q9UL39;  
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
 DT 01-NOV-1999, sequence version 1.  
 DT 07-FEB-2006, entry version 27.  
 DE Heparanase precursor (EC 3.2.-.-) (Heparanase-1) (Hpa1) (Endo-  
 DE glucuronidase) [Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa  
 DE subunit].  
 GN Name:HPSP; Synonyms:HEP, HPA, HPA1, HPR1, HPSE1, HSE1;  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.  
 RC TISSUE=Placenta;  
 RX MEDLINE=99335379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;  
 RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C.,  
 RA Seddon A.P., Giorgio N.A., Bohlen P.;  
 RT "Cloning and functional expression of a human heparanase gene.";   
 RL Biochem. Biophys. Res. Commun. 261:183-187(1999).  
 [2]  
 RN NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN  
 RP SEQUENCE OF 158-168; 326-337 AND 447-491.  
 RC TISSUE=Embryonic fibroblast;  
 RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;  
 RA Toyoshima M., Nakajima M.;  
 RT "Human heparanase. Purification, characterization, cloning, and  
 RT expression.";   
 RL J. Biol. Chem. 274:24153-24160(1999).  
 [3]  
 RN NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.  
 RP PubMed=10395325; DOI=10.1038/10518;  
 RX Vlodavsky I., Friedmann Y., Elkin M., Aingorn H., Atznon R.,  
 RA Ighai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,  
 RA Spector L., Pecker I.;  
 RT "Mammalian heparanase: gene cloning, expression and function in tumor  
 RT progression and metastasis.";   
 RL Nat. Med. 5:793-802(1999).  
 [4]  
 RN NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE  
 RP OF 158-174; 263-272; 326-337; 433-436; 438-443; 466-468 AND 478-483.  
 RC TISSUE=Placenta;  
 RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;  
 RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,

Parish C.R.;  
 RT "Cloning of mammalian heparanase, an important enzyme in tumor  
 RT invasion and metastasis.";   
 RL Nat. Med. 5:803-809(1999).  
 [5]  
 RN NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE  
 RP SPECIFICITY.  
 RC TISSUE=Placenta;  
 RX MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;  
 RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;  
 RT "Heparanase expression in invasive trophoblasts and acute vascular  
 RT damage.";   
 RL Glycobiology 10:467-475(2000).  
 [6]  
 RN NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.  
 RP PubMed=11547900; DOI=10.1023/A:1011375624902;  
 RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O.,  
 RA Aviv A., Elkin M., Pecker I., Peretz T., Vlodavsky I.;  
 RT "Molecular properties and involvement of heparanase in cancer  
 RT progression and mammary gland morphogenesis.";   
 RL J. Mammary Gland Biol. Neoplasia 6:311-322(2001).  
 [7]  
 RN NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163,  
 RP SUBUNITS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.  
 RC TISSUE=Placenta;  
 RX PubMed=12713442; DOI=10.1042/BJ20030318;  
 RA McKenzie E., Young K., Hircok M., Bennett J., Bhaman M., Felix R.,  
 RA Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S.,  
 RA Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M.,  
 RA Parekh R., Stubberfield C.;  
 RT "Biochemical characterization of the active heterodimer form of human  
 RT heparanase (Hpa1) protein expressed in insect cells.";   
 RL Biochem. J. 373:423-435(2003).  
 [8]  
 RN NUCLEOTIDE SEQUENCE [MRNA].  
 RP Pinhal M.A., Semedo P.;  
 RT "Cloned heparanase from MCF-7 cells.";   
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.  
 [9]  
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
 RP TISSUE=Small intestine;  
 RC Tanaka A., Yokoyama S.;  
 RT Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.  
 [10]  
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
 RP TISSUE=Pancreas;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,  
 RA Diatchenko M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Baha S., Locuelli N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";   
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 [11]  
 RN MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.  
 RP PubMed=11123890; DOI=10.1021/bi002080p;  
 RA Hulett M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,  
 RA Gready J.E., Parish C.R.;

RT "Identification of active-site residues of the pro-metastatic  
 RL endoglycosidase heparanase."; [2000].  
 RN [12]  
 RP N-GLYCOSYLATION, AND MUTAGENESIS OF ASN-162; ASN-178; ASN-200;  
 RX ASN-217; ASN-238 AND ASN-459.  
 RA PubMed=14573609; DOI=10.1074/jbc.M300541200;  
 RL Simizu S., Ishida K., Wierzbicka M.K., Osada H.;  
 RT "Secretion of heparanase protein is regulated by glycosylation in  
 RT human tumor cell lines."; [2004].  
 RL J. Biol. Chem. 279:2697-2703 (2004).  
 RN [13]  
 RP SUBCELLULAR LOCATION.  
 RX PubMed=15292202; DOI=10.1074/jbc.M402131200;  
 RA Gingis-Velitski S., Zetser A., Kaplan V., Ben-Zaken O., Cohen E.,  
 RL Levy-Adam F., Bashenko Y., Flugelman M.Y., Vlodavsky I., Ilan N.;  
 RT "Heparanase uptake is mediated by cell membrane heparan sulfate  
 RT proteoglycans."; [2005].  
 RL J. Biol. Chem. 279:44084-44092 (2004).  
 RN [14]  
 RP BIOPHYSICOCHEMICAL PROPERTIES, PROCESSING, AND SUBCELLULAR LOCATION.  
 RX PubMed=15848168; DOI=10.1016/j.febslet.2005.03.030;  
 RA Cohen E., Atzmon R., Vlodavsky I., Ilan N.;  
 RL "Heparanase processing by lysosomal/endosomal protein preparation."; [2005].  
 RN [15]  
 RP SUBCELLULAR LOCATION, PROCESSING, AND MUTAGENESIS OF TYR-156.  
 RX PubMed=15659389; DOI=10.1074/jbc.M413370200;  
 RA Aboud-Jarrou G., Rangini-Guetta Z., Angorn H., Atzmon R.,  
 RL Elgavish S., Peretz T., Vlodavsky I.;  
 RT "Site-directed mutagenesis, proteolytic cleavage, and activation of  
 RT human heparanase."; [2005].  
 RL J. Biol. Chem. 280:13568-13575 (2005).  
 RN [16]  
 RP DOMAINS, AND MUTAGENESIS OF LYS-158 AND LYS-161.  
 RX PubMed=15760902; DOI=10.1074/jbc.M41454200;  
 RA Levy-Adam F., Aboud-Jarrou G., Guerrini M., Beccati D.,  
 RL Vlodavsky I., Ilan N.;  
 RT "Identification and characterization of heparin/heparan sulfate  
 RT binding domains of the endoglycosidase heparanase."; [2005].  
 RL J. Biol. Chem. 280:20457-20466 (2005).  
 RN [17]  
 RP VARIANT SER-260.  
 RX PubMed=15334672;  
 RA Chen X.P., Liu Y.B., Rui J., Peng S.Y., Peng C.H., Zhou Z.Y.,  
 RL Shi L.H., Shen H.W., Xu B.;  
 RT "Heparanase mRNA expression and point mutation in hepatocellular  
 RT carcinoma."; [2004].  
 RL World J. Gastroenterol. 10:2795-2799 (2004).  
 CC -1- FUNCTION: Endoglycosidase which is a cell surface and  
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
 CC proteoglycans. Also implicated in the extravasation of leukocytes  
 CC and tumor cell lines. Due to its contribution to metastasis and  
 CC cancer therapies, it is considered to be a potential target for anti-  
 CC cancer therapies.  
 CC -1- ENZYME REGULATION: Inhibited by EDTA, laminarin sulfate and, to a  
 CC lower extent, by heparin and sulfamin and activated by calcium and  
 CC magnesium (By similarity).  
 CC -1- BIOPHYSICOCHEMICAL PROPERTIES:  
 CC pH dependence:  
 CC Optimum pH is 4-6;  
 CC -1- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
 CC subunits, the proteolytic products.  
 CC -1- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.  
 CC Secreted, internalised and transferred to late endosomes/lysosomes  
 CC as a proheparanase. In lysosomes, it is processed into the active  
 CC form, the heparanase. The uptake or internalisation of  
 CC proheparanase is mediated by HSPGs. Heparin appears to be a  
 CC competitor and retain proheparanase in the extracellular medium.  
 CC -1- TISSUE SPECIFICITY: Highly expressed in placenta and spleen and  
 CC weakly expressed in lymph node, thymus, peripheral blood  
 CC leukocytes, bone marrow, endothelial cells, fetal liver and tumor

CC tissues.  
 CC -1- PTM: Proteolytically processed. The cleavage of the 65 kDa form  
 CC leads to the generation of a linker peptide, 8 kDa and 50 kDa  
 CC product. The active form, the 8/50 kDa heterodimer, is resistant  
 CC to degradation. Complete removal of the linker peptide appears to  
 CC be a prerequisite to the complete activation of the enzyme.  
 CC -1- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears  
 CC to be essential for its solubility.  
 Query Match 100.0%; Score 82; DB 1; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 SWELGNEPNSFLKKA 15  
 DB 219 SWELGNEPNSFLKKA 233  
 RESULT 2  
 HPSE\_MOUSE  
 ID HPSE\_MOUSE STANDARD; PRT; 535 AA.  
 AC Q6YG21; Q8K3K3;  
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
 DT 11-OCT-2005, sequence version 2.  
 DT 07-MAR-2006, entry version 13.  
 DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) [Contains:  
 DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].  
 GN Name=Hspe; Synonyms=Hpa;  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA].  
 RC STRAIN=JUL/J; TISSUE=Spleen;  
 RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;  
 RA Hulet M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,  
 RA Parish C.R.;  
 RT "Cloning of mammalian heparanase, an important enzyme in tumor  
 RT invasion and metastasis";  
 RL Nat. Med. 5:803-809 (1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 28-57 AND 150-179,  
 RP GLYCOSYLATION, BIOPHYSICOCHEMICAL PROPERTIES, ENZYME REGULATION, AND  
 RP SUBUNITS.  
 RC STRAIN=FVB; TISSUE=Embryo;  
 RX MEDLINE=22481152; PubMed=12837765; DOI=10.1074/jbc.M300925200;  
 RA Miao H.-Q., Navarro E., Patel S., Sargent D., Koo H., Wan H.,  
 RA Plata A., Zhou Q., Ludwig D., Bohlen P., Kussie P.;  
 RT "Cloning, expression, and purification of mouse heparanase."; [2002].  
 RL Protein Expr. Purif. 26:425-431 (2002).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.  
 RX MEDLINE=22841152; PubMed=12837765; DOI=10.1074/jbc.M300925200;  
 RA Gong F., Jemth P., Galvis M.L.E., Vlodavsky I., Horner A., Lindahl U.,  
 RA Li J.-P.;  
 RT "Processing of macromolecular heparin by heparanase."; [2003].  
 RL J. Biol. Chem. 278:35152-35158 (2003).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
 RC STRAIN=C57BL/6J, and NOD; TISSUE=Thymus;  
 RX PubMed=16141072; DOI=10.1126/science.11112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Wilm M., Wang L.G., Aidinis V., Allen J.E.,  
 RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
 RA Banal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,

RA Georgii-Hemming P., Gingers T.R., Gojobori T., Green R.E.,  
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Huninicki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kanpin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavesi G., Pesole G.,  
RA Petrovsky N., Piazza S., Reid J.J., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.B., Sanderlin A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,  
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Nimomiya N.,  
RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayaishizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563(2005).  
CC -!- FUNCTION: Endoglycosidase which is a cell surface and  
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
CC proteoglycans. Also implicated in the extravasation of leukocytes  
CC and tumor cell lines. Contributes to metastasis and angiogenesis  
CC (By similarity).  
CC  
CC -!- ENZYME REGULATION: Inhibited by EDTA and activated by calcium and  
CC magnesium (By similarity). Inhibited by laminarin sulfate and, to  
CC a lower extent, by heparin and sulfamin.  
CC -!- BIOPHYSICOCHEMICAL PROPERTIES:  
CC pH dependence:  
CC Optimum pH is 5;  
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
CC subunits, the proteolytic products.  
CC -!- SUBCELLULAR LOCATION: Lysosomal, membrane-associated and secreted.  
CC Secreted, internalised and transferred to late endosomes/lysosomes  
CC as a proheparanase. In lysosomes, it is processed into the active  
CC form, the heparanase. The uptake or internalisation of  
CC proheparanase is mediated by HSPGs. Heparin appears to be a  
CC competitor and retain proheparanase in the extracellular medium  
CC (By similarity).  
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form  
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa  
CC product. The active form, the 8/50 kDa heterodimer, is resistant  
CC to degradation. Complete removal of the linker peptide appears to  
CC be a prerequisite to the complete activation of the enzyme (By  
CC similarity).  
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears  
CC to be essential for its solubility.  
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC DR EMBL; AF359507; AAQ15188.1; -; mRNA.  
CC DR EMBL; AY077467; AAL76083.1; -; mRNA.  
CC DR EMBL; AY151051; AAN41636.1; -; mRNA.  
CC DR EMBL; AK040471; BAC30600.1; -; mRNA.  
CC DR EMBL; AK154628; BA32725.1; -; mRNA.  
CC DR Ensembl; ENSMUSG00000035273; Mus musculus.  
CC DR MGI; MGI:1343124; Hspe.  
CC DR GO; GO:0005578; C:extracellular matrix (sensu Metazoa); TAS.  
CC DR InterPro; IPR005199; Glyco\_hydro\_79\_N.  
CC Pfam; PF03662; Glyco\_hydro\_79n; 1.

KW Calcium; Direct protein sequencing; Glycoprotein; Hydrolase; Lysosome;  
KW Magnesium; Membrane; Signal.  
FT SIGNAL 1 27 By similarity.  
FT CHAIN 28 101 Heparanase 8 kDa subunit.  
FT FT /FTID=PRO\_0000042263.  
FT FT Linker peptide (By similarity).  
FT FT /FTID=PRO\_0000042264.  
FT FT Heparanase 50 kDa subunit.  
FT FT /FTID=PRO\_0000042265.  
FT FT Heparin/Hs-binding (By similarity).  
FT FT Heparin/Hs-binding (By similarity).  
FT FT Proton donor (Potential).  
FT FT Nucleophile (Potential).  
FT FT N-linked (GlcNAc. . .) (Potential).  
FT FT N-linked (GlcNAc. . .) (Potential).  
FT FT N-linked (GlcNAc. . .) (Potential).  
FT FT N-linked (GlcNAc. . .) (Potential).  
FT FT N-linked (GlcNAc. . .) (Potential).  
FT FT N-linked (GlcNAc. . .) (Potential).  
FT FT K -> R (in Ref. 3).  
FT FT W -> S (in Ref. 3).  
FT FT NGS -> DGL (in Ref. 1, 2 and 4).  
FT FT E -> K (in Ref. 3).  
FT FT G -> A (in Ref. 3).  
FT FT Y -> H (in Ref. 1, 2 and 4).  
FT FT V -> I (in Ref. 1, 2 and 4).  
FT FT SEQUENCE 531 531  
SQ SEQUENCE 535 AA; 60050 MW; AF19B28B7CD03F7B CRC64;  
Query Match 92.7%; Score 76; DB 1; Length 535;  
Best Local Similarity 93.3%; Pred. No. 0.00028;  
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 SWELGNPNSFLKKA 15  
|||||||  
Db 211 SWELGNPNSFWKKA 225  
RESULT 3  
ID\_HPSE RAT STANDARD; PRT; 536 AA.  
AC Q71RP1; Q9QZP8;  
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
DT 05-JUL-2004, sequence version 1.  
DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) [Contains:  
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].  
GN Name=Hspe; Synonyms=Hep;  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridea; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA].  
RC TISSUE=Placenta;  
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;  
RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,  
RA Parish C.R.;  
RT "Cloning of mammalian heparanase, an important enzyme in tumor  
RT invasion and metastasis.";  
RL Nat. Med. 5:803-809(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.  
RX MEDLINE=22194309; PubMed=12077130; DOI=10.1074/jbc.M203282200;  
RA Podyma-Inoue K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;  
RT "Characterization of heparanase from a rat parathyroid cell line.";  
RL J. Biol. Chem. 277:32459-32465(2002).  
CC -!- FUNCTION: Endoglycosidase which is a cell surface and  
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
CC proteoglycans. Also implicated in the extravasation of leukocytes  
CC and tumor cell lines. Contributes to metastasis and angiogenesis  
CC (By similarity).  
CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower

NCBI\_TaxID=134510;  
 [1]  
 RN NUCLEOTIDE SEQUENCE.  
 RP TISSUE=Kidney;  
 RC Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
 RA "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
 RT cloning and identification of a novel splice variant.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).  
 [2]  
 RN NUCLEOTIDE SEQUENCE.  
 RP TISSUE=Kidney;  
 RC Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
 RA "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
 RT cloning and identification of a novel splice variant.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 -----  
 DR ENBL; AM085494; CAJ30021.1; -; mRNA.  
 SQ SEQUENCE 558 AA; 62737 MW; 07BAF8F55849EEB7 CRC64;  
  
 Query Match 92.7%; Score 76; DB 2; Length 558;  
 Best Local Similarity 93.3%; Pred. No. 0.00029;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 1 SWELGNPNPSFLKKA 15  
 |||||  
 DB 250 SWELGNPNPSFWKKA 264  
  
 RESULT 5  
 Q333X6 SPAJJD  
 ID Q333X6 SPAJJD PRELIMINARY; PRT; 574 AA.  
 AC Q333X6;  
 DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.  
 DT 06-DEC-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 3.  
 DE Heparanase.  
 DE  
 GN Spalax judeae (Blind subterranean mole rat).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muroidae; Spalacidae; Spalacinae; Spalax.  
 OX NCBI\_TaxID=134510;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Kidney;  
 RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
 RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
 RT cloning and identification of a novel splice variant.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).  
 [2]  
 RN NUCLEOTIDE SEQUENCE.  
 RP TISSUE=Kidney;  
 RC Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
 RA "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
 RT cloning and identification of a novel splice variant.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 -----  
 DR ENBL; AM085493; CAJ30020.1; -; mRNA.  
 SQ SEQUENCE 574 AA; 64515 MW; 3AE8B13F07451684 CRC64;  
  
 Query Match 92.7%; Score 76; DB 2; Length 574;  
 Best Local Similarity 93.3%; Pred. No. 0.0003;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 1 SWELGNPNPSFLKKA 15  
 |||||  
 DB 250 SWELGNPNPSFWKKA 264

```

Query Match      92.7%; Score 76; DB 2; Length 574;
Best Local Similarity 93.3%; Pred. No. 0.0003;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
DB 250 SWELGNEPNSFWKKA 264
|||||

RESULT 6
Q333X7_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X7;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax carmeli.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164324;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Aviavi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085492; CAJ30019.1; -; mRNA.
SQ SEQUENCE 574 AA; 64459 MW; 9FD19DCBABB99DE CRC64;

Query Match      92.7%; Score 76; DB 2; Length 574;
Best Local Similarity 93.3%; Pred. No. 0.0003;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
DB 250 SWELGNEPNSFWKKA 264
|||||

RESULT 7
Q333X8_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X8;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax golani.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=191382;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Aviavi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085491; CAJ30018.1; -; mRNA.
SQ SEQUENCE 574 AA; 64555 MW; 40EBBEPEC7D0BCB34 CRC64;

Query Match      92.7%; Score 76; DB 2; Length 574;
Best Local Similarity 93.3%; Pred. No. 0.0003;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
DB 250 SWELGNEPNSFWKKA 264
|||||

RESULT 8
Q333X9_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X9;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax galili.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164323;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Aviavi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085490; CAJ30017.1; -; mRNA.
SQ SEQUENCE 574 AA; 64525 MW; 1635865051B380D0 CRC64;

Query Match      92.7%; Score 76; DB 2; Length 574;
Best Local Similarity 93.3%; Pred. No. 0.0003;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
DB 250 SWELGNEPNSFWKKA 264
|||||

RESULT 9
HPSE CHICK STANDARD; PRT; 523 AA.
AC Q90VK5;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-DEC-2001, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Heparanase precursor (EC 3.2.-.-).
GN Name=HPSE; Synonyms=HPA;
OS Gallus gallus (Chicken)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RX MEDLINE=21369959; PubMed=11387326; DOI=10.1074/jbc.M102462200;
RA Goldshmidt O., Zcharia E., Aingorn H., Guatta-Rangini Z., Atzmon R.,
RA Michal I., Pecker I., Mitrani E., Vlodavsky I.;
```



RT "Expression pattern and secretion of human and chicken heparanase are  
 determined by their signal peptide sequence.";  
 RL J. Biol. Chem. 276:29178-29187(2001).  
 CC -1- FUNCTION: Endoglycosidase which is a cell surface and  
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
 CC proteoglycans (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted  
 CC (By similarity).  
 CC -1- PTM: N-glycosylated (By similarity).  
 CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL: AY037007; AAK82648.1; -; mRNA.  
 DR EMBL: ENSGALG0000011203; Gallus gallus.  
 DR InterPro: IPR005199; Glyco\_hydro\_79\_N.  
 DR Pfam: PF03662; Glyco\_hydro\_79n; 1.  
 DR Glycoprotein; Hydrolase; Lysosome; Membrane; Signal.  
 KW SIGNAL 1 18 Potential.  
 FT CHAIN 19 523 Heparanase.  
 FT REGION 137 141 /FTid=PRO\_0000042259.  
 FT REGION 250 260 Heparin/HS-binding (By similarity).  
 FT ACT\_SITE 204 204 Heparin/HS-binding (By similarity).  
 FT ACT\_SITE 323 323 Proton donor (Potential).  
 FT CARBOHYD 141 141 Nucleophile (Potential).  
 FT CARBOHYD 196 196 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 436 436 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 439 439 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 523 523 N-linked (GlcNAc...) (Potential).  
 SQ SEQUENCE 523 AA; 58386 MW; 8BE0B7B18C9BF881 CRC64;  
 Query Match 89.0%; Score 73; DB 1; Length 523;  
 Best Local Similarity 86.7%; Pred. No. 0.00092;  
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 SWELGNPNPNSFLKKA 15  
 |||||  
 DB 198 SWELGNPNPNSFRKKS 212  
 |||||  
 RESULT 10  
 HPSE BOVIN  
 ID HPSE\_BOVIN STANDARD; PRT; 545 AA.  
 AC Q9MY0;  
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
 DT 01-JUN-2001, sequence version 2.  
 DT 07-MAR-2006, entry version 15.  
 DE Heparanase precursor (EC 3.2.2.-) [Contains: Heparanase 8 kDa subunit;  
 DE Heparanase 50 kDa subunit].  
 GN Name=HPSE;  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
 OC Pecora; Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.  
 RC TISSUE=Placenta;  
 RX MEDLINE=21176669; PubMed=11277877;  
 RA Kizaki K., Nakano H., Nakano H., Takahashi T., Imai K., Hashizume K.;  
 RT "Expression of heparanase mRNA in bovine placenta during gestation.";  
 RL Reproduction 121:573-580(2001).  
 CC -1- FUNCTION: Endoglycosidase which is a cell surface and  
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
 CC proteoglycans. Also implicated in the extravasation of leukocytes  
 CC and tumor cell lines. Contributes to metastasis and angiogenesis  
 CC (By similarity).  
 CC -1- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower  
 CC extent, by heparin, sulfamin and EDTA. Activated by calcium and  
 CC magnesium (By similarity).  
 CC

CC -1- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
 CC subunits, the proteolytic products (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.  
 CC Secreted, internalised and transferred to late endosomes/lysosomes  
 CC as a proheparanase. In lysosomes, it is processed into the active  
 CC form, the heparanase. The uptake or internalisation of  
 CC proheparanase is mediated by HSPGs. Heparin appears to be a  
 CC competitor and retain proheparanase in the extracellular medium  
 CC (By similarity).  
 CC -1- TISSUE SPECIFICITY: Highly expressed in placenta and weakly in the  
 CC kidney, lung, spleen and uterus.  
 CC -1- PTM: Proteolytically processed. The cleavage of the 65 kDa form  
 CC leads to the generation of a linker peptide, 8 kDa and 50 kDa  
 CC product. The active form, the 8/50 kDa heterodimer, is resistant  
 CC to degradation. Complete removal of the linker peptide appears to  
 CC be a prerequisite to the complete activation of the enzyme (By  
 CC similarity).  
 CC -1- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears  
 CC to be essential for its solubility (By similarity).  
 CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL: AF281160; AAF87301.2; -; mRNA.  
 DR InterPro: IPR005199; Glyco\_hydro\_79\_N.  
 DR Pfam: PF03662; Glyco\_hydro\_79n; 1.  
 DR Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;  
 KW SIGNAL 1 37 By similarity.  
 FT CHAIN 38 111 Heparanase 8 kDa subunit (By similarity).  
 FT PROPEP 112 159 Linker peptide.  
 FT CHAIN 160 545 Heparanase 50 kDa subunit (By  
 FT similarity).  
 FT REGION 160 164 /FTid=PRO\_0000042258.  
 FT REGION 272 282 Heparin/HS-binding (Potential).  
 FT ACT\_SITE 227 227 Proton donor (Potential).  
 FT ACT\_SITE 345 345 Nucleophile (Potential).  
 FT CARBOHYD 164 164 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 219 219 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 461 461 N-linked (GlcNAc...) (Potential).  
 SQ SEQUENCE 545 AA; 61077 MW; FAC4BDFDF855B933 CRC64;  
 Query Match 89.0%; Score 73; DB 1; Length 545;  
 Best Local Similarity 86.7%; Pred. No. 0.00097;  
 Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 SWELGNPNPNSFLKKA 15  
 |||||  
 DB 221 SWELGNPNPNSFORKA 235  
 |||||  
 RESULT 11  
 Q4TGC8\_TETNG  
 ID Q4TGC8\_TETNG PRELIMINARY; PRT; 255 AA.  
 AC Q4TGC8;  
 DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
 DT 19-JUL-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 4.  
 DE Chromosome undetermined SCAF3783, whole genome shotgun sequence.  
 DE (Fragment).  
 GN ORFNames=GSTENG000001168001;  
 OS Tetraodon nigroviridis (Green puffer).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontidae; Tetraodontidae; Tetraodon.  
 OX NCBI\_TaxID=99883;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.



```

RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Croliius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RG NUCLEOTIDE SEQUENCE.
RL Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAEA01003783; CAF88054.1; -; Genomic_DNA.
FT NON_TER 1
FT NON_TER 255
FT SEQUENCE 255 AA; 28562 MW; 07F542A9C755E3F0 CRC64;
SQ
Query Match 87.8%; Score 72; DB 2; Length 255;
Best Local Similarity 86.7%; Pred. No. 0.0062;
Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
Db 54 SWELGNEPNSYEKKA 58
|||||: |||
|||||: |||

RESULT 12
Q4SYF6_TETNG PRELIMINARY; PRT; 533 AA.
AC
Q4SYF6;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Chromosome undetermined SCAF12073, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG00010356001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontoidea; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Croliius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
```

```

RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RG NUCLEOTIDE SEQUENCE.
RL Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAEA01012073; CAF94326.1; -; Genomic_DNA.
FT NON_TER 1
FT NON_TER 533
FT SEQUENCE 533 AA; 60100 MW; 9B00A7C8780100FF CRC64;
SQ
Query Match 87.8%; Score 72; DB 2; Length 533;
Best Local Similarity 86.7%; Pred. No. 0.0014;
Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SWELGNEPNSFLKKA 15
Db 174 SWELGNEPNSYEKKA 188
|||||: |||
|||||: |||

RESULT 13
HPSE2_HUMAN
ID HPSE2_HUMAN STANDARD; PRT; 592 AA.
AC Q8WQ2; Q5VUH4; Q5VUH5; Q5VUH6; Q8WQ1; Q9HB37; Q9HB38; Q9HB39;
DT 25-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 25-OCT-2005, sequence version 2.
DT 07-MAR-2006, entry version 16.
DE Heparanase-2 (EC 3.2.-.-) (Hpa2).
GN Name=HPSE2; Synonyms=HPA2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA] (ISOFORMS 1; 3 AND 4), TISSUE SPECIFICITY,
RP AND SUBCELLULAR LOCATION.
RC TISSUE=Heart;
RX MEDLINE=20483645; PubMed=11027606; DOI=10.1006/bbrc.2000.3586;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hircock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;
RT "Cloning and expression profiling of Hpa2, a novel mammalian
RT heparanase family member.";
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA] (ISOFORMS 1 AND 2).
RC TISSUE=Prostate;
RA Legoux P., Legoux R., O'Brien D., Salome M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA], AND VARIANT TYR-579.
RX PubMed=15164054; DOI=10.1038/nature02462;
RA Deloukas P., Earthworm M.E., Grafham D.V., Rubinfeld M., French L.,
RA Steward C.A., Sims S.K., Jones M.C., Searle S., Scott C., Howe K.,
RA Hunt S.E., Andrews T.D., Gilbert J.G.R., Swarbreck D., Ashurst J.L.,
RA Taylor A., Battles J., Bird C.P., Ainscough R., Almeida J.P.,
RA Ashwell R.I.S., Ambrose K.D., Babbage A.K., Baggeley C.L., Bailey J.,
RA Brown J.R., Bates K., Beasley H., Bray-Allen S., Brown A.J.,
RA Brown J.Y., Burford D.C., Burrill W., Burroll J., Cahill P., Camire D.,
RA Carter N.P., Chapman J.C., Clark S.Y., Clarke G., Clee C.M., Clegg S.,
RA Corby N., Coulson A., Dhani P., Dutta I., Dunn M., Faulkner L.,
RA Frankish A., Frankland J.A., Garner P., Garnett J., Gribble S.,
RA Griffiths C., Grocock R., Gustafson E., Hammond S., Harley J.L.,
RA Hart E., Heath P.D., Ho T.P., Hopkins B., Horne J., Howden P.,
RA Huckle E., Hynds C., Johnson C., Johnson D., Kana A., Kay M.,
RA Kimberley A.M., Kershaw J.K., Kokkinaki M., Laird G.K., Lawlor S.,
```



```

Best Local Similarity 81.8%; Pred. No. 0.47;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSF 11
    |||||:::
Db 256 SWELGNEPNY 266

RESULT 14
Q2M1H9 HUMAN
ID Q2M1H9 HUMAN PRELIMINARY; PRT; 592 AA.
AC Q2M1H9;
DT 21-FEB-2006, integrated into UniProtKB/TrEMBL.
DT 21-FEB-2006, sequence version 1.
DT 21-FEB-2006, entry version 1.
DE Heparanase 2.
GN Name=HPS2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=PCR rescued clones;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Matra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences".
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=PCR rescued clones;
RG NIH MGC Project;
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; BC112356; AAI12357.1; -; mRNA.
SQ SEQUENCE 592 AA; 66610 MW; 94689E1C2A74359F CRC64;

Query Match 70.7%; Score 58; DB 2; Length 592;
Best Local Similarity 81.8%; Pred. No. 0.47;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSF 11
    |||||:::
Db 256 SWELGNEPNY 266

RESULT 15
Q4TB80 TETNG
ID Q4TB80 TETNG PRELIMINARY; PRT; 597 AA.
AC Q4TB80;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.

```

```

DT 07-FEB-2006, entry version 4.
DE Chromosome 17 SCAF7180, whole genome shotgun sequence. (Fragment).
GN ORFNames=GSTENG00003868001;
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jallou O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Bionet C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Crnaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype."
RL Nature 431:946-957 (2004).
[2]
RN NUCLEOTIDE SEQUENCE.
RP Genoscope; Whitehead Institute Centre for Genome Research;
RG Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
RL CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; CAAB01007180; CAP89852.1; -; Genomic_DNA.
FT NON TER. 597 597
SQ SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;

Query Match 70.7%; Score 58; DB 2; Length 597;
Best Local Similarity 81.8%; Pred. No. 0.47;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SWELGNEPNSF 11
    |||||:::
Db 256 SWELGNEPNY 268

RESULT 16
Q8T108 BOMMO
ID Q8T108 BOMMO PRELIMINARY; PRT; 515 AA.
AC Q8T108;
DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2002, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Heparanase-like protein.
GN Name=Bmhepa;
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
OC Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=p50; TISSUE=Posterior silk gland;
RA Koike Y., Mita K., Suzuki M.G., Maeda S., Abe H., Osoegawa K.,
RA deJong P.J., Shimada T.;
RT "Genomic sequence of a 320-kb segment of the Z chromosome of Bombyx
RT mori containing a kettin ortholog."
RL Mol. Genet. Genomics 269:137-149 (2003).

```

```
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AB079860; BAB85191.1; -; Genomic DNA.
CC EMBL; AB090307; BAC10612.1; -; Genomic DNA.
CC InterPro; IPR005199; Glyco_hydro_79_N.
CC Pfam; PF03662; Glyco_hydro_79n; 1.
CC SEQUENCE 515 AA; 59770 MW; FB8100ABBEEDDADB CRC64;

Query Match      59.0%; Score 57; DB 2; Length 515;
Best Local Similarity 90.0%; Pred. No. 0.6;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 WELGNEPNSF 11
   |:|||||
Db 194 WQLGNEPNSF 203

RESULT 17
BGLR_CANFA
ID BGLR_CANFA STANDARD; PRT; 651 AA.
AC O18835;
DT 15-JUL-1998, integrated into UniProtKB/Swiss-Prot.
DT 01-JAN-1998, sequence version 1.
DT 21-FEB-2006, entry version 40.
DE Beta-glucuronidase precursor (EC 3.2.1.31).
GN Name=GUSB;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC Canis.
OX NCBI_TaxID=9615;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND VARIANT MPS VII HLS-166.
RA MEDLINE=98190525; PubMed=9521879; DOI=10.1006/geno.1997.5189;
RA Ray J., Bouvet A., Desanto C., Fyfe J.C., Xu D., Wolfe J.H.,
RA Aguirre G.D., Patterson D.F., Haakins M.E., Henthorn P.S.;
RT "Cloning of the canine beta-glucuronidase cDNA, mutation
RT identification in canine MPS VII, and retroviral vector-mediated
RT correction of MPS VII cells.";
RL Genomics 48:248-253(1998).
CC -1- FUNCTION: Plays an important role in the degradation of dermatan
CC and keratan sulfates [By similarity].
CC -1- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = D-glucuronate
CC + an alcohol.
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SUBCELLULAR LOCATION: Lysosome.
CC -1- DISEASE: Defects in GUSB are the cause of mucopolysaccharidosis
CC type VII (MPS VII), an inherited disease reported in humans, mice,
CC cats, and dogs.
CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 2 family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AF019759; AAC48809.1; -; mRNA.
CC HSP; P08236; 1BHG.
CC SMR; O18835; 22-631.
CC Ensembl; ENSCAF00000010193; Canis familiaris.
CC InterPro; IPR006101; Glyco_hydro_2.
CC InterPro; IPR006102; Glyco_hydro_2IG.
CC InterPro; IPR006103; Glyco_hydro_2TIM.
CC InterPro; IPR006104; Glyco_hydro_S_bd.
CC Pfam; PF00703; Glyco_hydro_2; 1.
CC Pfam; PF02836; Glyco_hydro_2_C; 1.
CC Pfam; PF02837; Glyco_hydro_2_N; 1.
CC PRINTS; PR00132; GLHYDLASE2.
CC PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
CC PROSITE; PS00608; GLYCOSYL_HYDROL_F2_1.
CC Disease mutation; Glycoprotein; Glycosidase; Hydrolase; Lysosome;
CC Signal.
CC SIGNAL 1 22 By similarity.
FT
```

```
FT CHAIN 23 651 Beta-glucuronidase.
FT ACT_SITE 450 450 Proton donor (By similarity).
FT CARBOHYD 172 172 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 419 419 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 630 630 N-linked (GlcNAc...) (Potential).
FT VARIANT 166 166 R -> H (in MPS VII; loss of activity).
SQ SEQUENCE 651 AA; 74433 MW; E8991B1E65C60120 CRC64;

Query Match      63.4%; Score 52; DB 1; Length 651;
Best Local Similarity 64.3%; Pred. No. 5.9;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKKA 15
   |:|||||
Db 445 WSVANEPTSFLKPA 458

RESULT 18
Q4FAT7_PIG
ID Q4FAT7_PIG PRELIMINARY; PRT; 652 AA.
AC Q4FAT7;
DT 30-AUG-2005, integrated into UniProtKB/TrEMBL.
DT 30-AUG-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Beta-glucuronidase.
GN Name=GUSB;
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
OC Sus.
OX NCBI_TaxID=9823;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Beck J., Knorr C., Brenig B.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = D-glucuronate
CC + an alcohol.
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; DQ095863; AA203639.1; -; Genomic DNA.
CC SMR; Q4FAT7; 24-632.
CC GO; GO:0005764; C:lysosome; IEA.
CC GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl...; IEA.
CC GO; GO:0005975; P:carbohydrate metabolism; IEA.
CC InterPro; IPR006101; Glyco_hydro_2.
CC Pfam; PF00703; Glyco_hydro_2; 1.
CC Pfam; PF02836; Glyco_hydro_2_C; 1.
CC Pfam; PF02837; Glyco_hydro_2_N; 1.
CC PRINTS; PR00132; GLHYDLASE2.
CC PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
CC PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
CC Lysosome.
CC SEQUENCE 652 AA; 74710 MW; 7B9386564DFE2CA7 CRC64;

Query Match      62.2%; Score 51; DB 2; Length 652;
Best Local Similarity 57.1%; Pred. No. 8.9;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKKA 15
   |:|||||
Db 446 WSVANEPTSFLKPA 459

RESULT 19
BGLR_FELCA
ID BGLR_FELCA STANDARD; PRT; 651 AA.
AC O97524;
DT 15-JUL-1999, integrated into UniProtKB/Swiss-Prot.
FT
```

DT 01-MAY-1999, sequence version 1.  
DT 21-FEB-2006, entry version 39.  
DE Beta-glucuronidase precursor (EC 3.2.1.31).  
GN Name=CUSB;  
OS Felis silvestris catus (Cat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Felidae;  
OC Felinae; Felis.  
OX NCBI\_TaxID=9685;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND VARIANT MPS VII LYS-351.  
RC TISSUE=Liver;  
RX MEDLINE=99296826; PubMed=10366443; DOI=10.1006/geno.1999.5825;  
RA Fyfe J.C., Kurzhals R.L., Lassealine M.E., Henthorn P.S., Alur P.R.,  
RA Wang P., Wolfe J.H., Giger U., Haskins M.E., Patterson D.F., Sun H.,  
RA Jain S., Yukki N.;  
RT "Molecular basis of feline beta-glucuronidase deficiency: an animal  
RT model of mucopolysaccharidosis VII.";  
RL Genomics 58:121-128(1999).  
CC -1- FUNCTION: Plays an important role in the degradation of dermatan  
CC and keratan sulfates.  
CC -1- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = D-glucuronate  
CC + an alcohol.  
CC -1- SUBUNIT: Homotetramer (By similarity).  
CC -1- SUBCELLULAR LOCATION: Lysosome.  
CC -1- DISEASE: Defects in GUSB are the cause of mucopolysaccharidosis  
CC type VII (MPS VII), an inherited disease reported in humans, mice,  
CC cats, and dogs.  
CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 2 family.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC ENBL; AF012423; AAD01498.1; -; mRNA.  
DR ENBL; AF012424; AAD01499.1; -; mRNA.  
DR HSSP; P08236; 1BHG.  
DR SNR; O97524; 22-631.  
DR InterPro; IPR006101; Glyco\_hydro\_2.  
DR InterPro; IPR006102; Glyco\_hydro\_2Ig.  
DR InterPro; IPR006103; Glyco\_hydro\_2Tm.  
DR InterPro; IPR006104; Glyco\_hydro\_2\_b0.  
DR Pfam; PF00703; Glyco\_hydro\_2; 1.  
DR Pfam; PF02836; Glyco\_hydro\_2\_C; 1.  
DR Pfam; PF02837; Glyco\_hydro\_2\_N; 1.  
DR PRINTS; PR00132; GLHYDRLASE2.  
DR PROSITE; PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
DR PROSITE; PS00608; GLYCOSYL\_HYDROL\_F2\_2; 1.  
DR Disease mutation; Glycoprotein; Glycosidase; Hydrolase; Lysosome;  
KW Signal.  
FT SIGNAL 1 22 By similarity.  
FT CHAIN 23 651 Beta-glucuronidase.  
FT /FTID=PRO\_0000012160.  
FT ACT\_SITE 450 450 Proton donor (By similarity).  
FT CARBOHYD 172 172 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 419 419 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 630 630 N-linked (GlcNAc...) (Potential).  
FT VARIANT 351 351 E -> K (in MPS VII).  
SQ SEQUENCE 651 AA; 74610 MW; 2AE30884B70D4232 CRC64;  
  
Query Match 61.0%; Score 50; DB 1; Length 651;  
Best Local Similarity 64.3%; Pred. No. 13;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 2 WEIENPNSFLKKA 15  
| : ||| ||| |  
Db 445 WSVANEPASFLKPA 458  
  
RESULT 20  
Q6FYV6 BARQU  
ID Q6FYV6 BARQU PRELIMINARY; PRT; 581 AA.  
AC Q6FYV6;  
DT 19-JUL-2004, integrated into UniProtKB/TrEMBL.

DT 19-JUL-2004, sequence version 1.  
DT 07-FEB-2006, entry version 8.  
DE Hypothetical protein.  
GN OrderedLocusNames=BQ10670;  
OS Bartonella quintana (Rochalimaea quintana).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Bartonellaceae; Bartonella.  
OX NCBI\_TaxID=803;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=Toulouse;  
RX PubMed=15210978; DOI=10.1073/pnas.0305659101;  
RA Almark U.C.M., Frank A.C., Karlberg E.O., Legault B.-A., Ardell D.H.,  
RA Canbaeck B., Eriksson A.-S., Naesslund A.K., Handley S.A., Huvet M.,  
RA La Scola B., Holmberg M., Andersson S.G.E.;  
RT "The louse-borne human pathogen Bartonella quintana is a genomic  
RT derivative of the zoonotic agent Bartonella henselae.";  
RL Proc. Natl. Acad. Sci. U.S.A. 101:9716-9721(2004).  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC ENBL; BX897700; CAF26534.1; -; Genomic DNA.  
DR Complete proteome; Hypothetical protein.  
SQ SEQUENCE 581 AA; 65072 MW; 146DEC3FF6339D24 CRC64;  
  
Query Match 59.8%; Score 49; DB 2; Length 581;  
Best Local Similarity 60.0%; Pred. No. 18;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
  
Qy 1 SWELGNPNPSFLKKA 15  
| : | | | | |  
Db 307 SWEVANHPTFSKLA 321  
  
RESULT 21  
Q2TYD4 ASPOR  
ID Q2TYD4 ASPOR PRELIMINARY; PRT; 260 AA.  
AC Q2TYD4;  
DT 24-JAN-2006, integrated into UniProtKB/TrEMBL.  
DT 24-JAN-2006, sequence version 1.  
DT 24-MAR-2006, entry version 3.  
DE Predicted protein.  
GN ORFNames=AO090103000292;  
OS Aspergillus oryzae.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5062;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=RIB 40;  
RX PubMed=16372010; DOI=10.1038/nature04300;  
RA Machida M., Asai K., Sano M., Tanaka T., Kumagai T., Terai G.,  
RA Kusumoto K., Arima T., Akita O., Kashiwagi Y., Abe K., Gomi K.,  
RA Horiuchi H., Kitamoto K., Kobayashi T., Takeuchi M., Denning D.W.,  
RA Galagan J.E., Nierman W.C., Yu J., Archer D.B., Bennett J.W.,  
RA Bhatnagar D., Cleveland T.E., Fedorova N.D., Gotoh O., Horikawa H.,  
RA Hosoyama A., Ichinomiya M., Igarashi R., Iwashita K., Juvvadi P.R.,  
RA Kato M., Kato Y., Kin T., Kokubun A., Maeda H., Maeyama N.,  
RA Maruyama J., Nagasaki H., Nakajima T., Oda K., Okada K., Paulsen I.,  
RA Sakamoto K., Sawano T., Takahashi M., Takase K., Terabayashi Y.,  
RA Wortman J.R., Yamada O., Yamagata Y., Anazawa H., Hata Y., Koide Y.,  
RA Komori T., Koyama Y., Minetoki T., Suharnan S., Tanaka A., Isono K.,  
RA Kuhara S., Ogasawara N., Kikuchi H.;  
RT "Genome sequencing and analysis of Aspergillus oryzae.";  
RL Nature 438:1157-1161(2005).  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC ENBL; AP007174; BAE65739.1; -; Genomic DNA.  
AC AP007174;  
SQ SEQUENCE 260 AA; 28377 MW; FA60BE90767F5982 CRC64;

```
Query Match      58.5%; Score 48; DB 2; Length 260;
Best Local Similarity 54.5%; Pred. No. 11;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy      1 SWELGNEPNSF 11
      :||:||||:
Db      168 AWEIGNEPDNY 178

RESULT 22
Q64Y35_BACFR PRELIMINARY; PRT; 435 AA.
JAC Q64Y35;
DT 25-OCT-2004, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 1.
DE Endo-1,4-beta-mannosidase.
GN OrderedLocusNames=BF0840;
OS Bacteroides fragilis.
OC Bacteria; Bacteroidetes; Bacteroidetes (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=817;
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC PubMed=15466707; DOI=10.1073/pnas.0404172101;
RA Kuwahara T., Yamashita A., Hirakawa H., Nakayama H., Okada N.,
RA Kuhara S., Hattori M., Hayashi T., Ohnishi Y.;
RT "Genomic analysis of Bacteroides fragilis reveals extensive DNA
RT inversions regulating cell surface adaptation.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14919-14924(2004).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
EMBL: AP006841; BAD47591.1; -; Genomic_DNA.
KW Complete proteome.
SQ SEQUENCE 435 AA; 48916 MW; 34D8423F8DD8E86CB CRC64;

Query Match      58.5%; Score 48; DB 2; Length 435;
Best Local Similarity 63.6%; Pred. No. 19;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy      1 SWELGNEPNSF 11
      :||:||||:
Db      215 SWQIGNEPRPF 225

RESULT 23
Q3IKG9_PSEHT PRELIMINARY; PRT; 1297 AA.
AC Q3IKG9;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 1.
DE Helicase, ATP-dependent (EC 3.6.1.15).
GN Name=hrpA; OrderedLocusNames=PSHA1144; ORFNames=PSHA1144;
OS Pseudalteromonas haloplanktis (strain TAC 125).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Pseudalteromonadaceae; Pseudalteromonas.
OX NCBI_TaxID=326442;
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=16169927; DOI=10.1101/gr.4126905;
RA Medigue C., Krin E., Pascal G., Barbe V., Bernsel A., Bertin P.N.,
RA Cheung F., Cruveiller S., D'Amico S., Duilio A., Fang G., Feller G.,
RA Ho C., Mangelon S., Marino G., Nilsson J., Parrilli E., Rocha E.P.C.,
RA Rouy Z., Sekowska A., Tutino M.L., Vallenet D., von Heijne G.,
RA Danchin A.;
RT "Coping with cold: the genome of the versatile marine Antarctic
RT bacterium Pseudalteromonas haloplanktis TAC125.";
RL Genome Res. 15:1325-1335(2005).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
```

```
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
EMBL: CR954246; CA186219.1; -; Genomic_DNA.
CC
DR GO:0005524; F:ATP binding; IEA.
DR GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO:0016787; F:hydrolase activity; IEA.
DR GO:0003676; F:nucleic acid binding; IEA.
DR GO:0017111; F:nucleoside-triphosphatase activity; IEA.
DR GO:0000166; F:nucleotide binding; IEA.
DR InterPro: IPR003593; AAA ATPase.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR011545; DEAD/DEAH_N.
DR InterPro: IPR010222; DEAH_box_HrpA.
DR InterPro: IPR011709; DUF1605.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR007502; Helicase_dom.
DR Pfam: PF0270; DEAD; 1.
DR Pfam: PF07717; DUF1605; 1.
DR Pfam: PF04408; HA2; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR SMART: SM00382; AAA; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR TIGRFAMs: TIGR01967; DEAH_box_HrpA; 1.
KW Complete proteome; Helicase; Hydrolase.
SQ SEQUENCE 1297 AA; 146786 MW; CED447350085FFLD CRC64;

Query Match      57.3%; Score 47; DB 2; Length 1297;
Best Local Similarity 53.8%; Pred. No. 98;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy      2 WELGNEPNSFLKK 14
      :|:||||:
Db      1019 WDFGELPNSYVKK 1031

RESULT 24
Q3CKA3_THEET PRELIMINARY; PRT; 627 AA.
ID Q3CKA3_THEET
AC Q3CKA3;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 22-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Glycoside hydrolase, family 5.
GN ORFNames=Teth39DRAFT_0967;
OS Thermoanaerobacter ethanolicus ATCC 33223.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=340099;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 33223;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Thermoanaerobacter
RT ethanolicus 39E.";
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 33223;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome of Thermoanaerobacter ethanolicus
RT 39E.";
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
```

```
CC -----
DR EMBL; AAKQ01000001; EAO65730.1; -; Genomic_DNA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
KW Hydrolase.
SQ SEQUENCE 627 AA; 72436 MW; BD4FB389872AB8FC CRC64;

Query Match 56.1%; Score 46; DB 2; Length 627;
Best Local Similarity 50.0%; Pred. No. 65;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLK 13
Db 139 WDLSPNDNVVK 150
|:| |||:|:|

RESULT 25
Q8RDH1 THETN PRELIMINARY; PRT; 637 AA.
AC Q8RDH1;
DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2002, sequence version 1.
DT 07-FEB-2006, entry version 18.
DE Hypothetical protein.
GN OrderedLocusNames=TT00061;
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=2192816; PubMed=11997336; DOI=10.1101/gr.219302;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of the T. tengcongensis genome.";
RL Genome Res. 12:1689-700(2002).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AEO12980; AAM23368.1; -; Genomic_DNA.
DR BIOCYC; ITEN119072:IT00061-MONOMER; -;
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00150; Cellulase; I.
KW Complete proteome.
SQ SEQUENCE 637 AA; 74206 MW; 619E35812AC83EBB CRC64;

Query Match 56.1%; Score 46; DB 2; Length 637;
Best Local Similarity 50.0%; Pred. No. 66;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLK 13
Db 149 WDLSPNDNVVK 160
|:| |||:|:|

RESULT 26
Q3E4N3 CHLAU PRELIMINARY; PRT; 651 AA.
AC Q3E4N3;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 1.
DT 07-FEB-2006, entry version 3.
DE Glycoside hydrolase, family 5.
GN ORFNames=CaurDRAFT_0879;
OS Chloroflexus aurantiacus J-10-fl.
OC Bacteria; Chloroflexi; Chloroflexales; Chloroflexaceae; Chloroflexus.
OX NCBI_TaxID=324602;
```

```
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=J-10-fl;
RC US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hamon N., Israni S., Pittluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Chloroflexus
RT aurantiacus J-10-fl.";
RL Submitted (SEP-2005) to the EMBL/GenBank/DBSJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=J-10-fl;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Chloroflexus aurantiacus
RT J-10-fl.";
RL Submitted (SEP-2005) to the EMBL/GenBank/DBSJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=J-10-fl;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hamon N., Israni S., Pittluck S., Richardson P.;
RL Submitted (SEP-2005) to the EMBL/GenBank/DBSJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AARH02000003; EAO60769.1; -; Genomic_DNA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
KW Hydrolase.
SQ SEQUENCE 651 AA; 71446 MW; 2737C9294AC882FE CRC64;

Query Match 56.1%; Score 46; DB 2; Length 651;
Best Local Similarity 77.8%; Pred. No. 68;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 WELGNEPNS 10
Db 350 WELGNEPDA 358
|:| |||:|:|

RESULT 27
Q2U2I3 ASPOR PRELIMINARY; PRT; 373 AA.
ID Q2U2I3 ASPOR PRELIMINARY; PRT; 373 AA.
AC Q2U2I3;
DT 24-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 24-JAN-2006, sequence version 1.
DT 07-MAR-2006, entry version 3.
DE Endo-beta-mannanase.
GN ORFNames=AO090038000444;
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=RIB 40;
RX PubMed=16372010; DOI=10.1038/nature04300;
RA Machida M., Asai K., Sano M., Tanaka T., Kumagai T., Terai G.,
RA Kusumoto K., Arima T., Akita O., Kashiwagi Y., Abe K., Gomi K.,
RA Horiuchi H., Kitamoto K., Kobayashi T., Takeuchi M., Denning D.W.,
RA Galagan J.E., Nierman W.C., Yu J., Archer D.B., Bennett J.W.,
```



Bhatnagar D., Cleveland T.E., Fedorova N.D., Gotoh O., Horikawa H., Hosoyana A., Ichinomiya M., Igarashi R., Iwashita K., Juvvadi P.R., Kato M., Kato Y., Kin I., Kokubun A., Maeda K., Maeyama N., Maruyama J., Nagasaki H., Nakajima T., Oda K., Okada K., Paulsen I., Sakamoto K., Sawano T., Takahashi M., Takase K., Terabayashi Y., Wortman J.R., Yamada O., Yamagata Y., Anazawa H., Hata Y., Koide Y., Komori T., Koyama Y., Minetoki T., Suharnan S., Tanaka A., Isono K., Kuhara S., Ogasawara N., Kikuchi H.,  
RT Nature 438:1157-1161(2005).  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
-----  
DR EMBL; AP007169; BAB64232.1; -; Genomic DNA.  
SQ SEQUENCE 373 AA; 41126 MW; 3AD40B2A6AFC7486 CRC64;  
  
Query Match 54.9%; Score 45; DB 2; Length 373;  
Best Local Similarity 87.5%; Pred. No. 55;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 SWELGNBP 8  
Db 190 AWELGNBP 197  
:|||||  
-----  
RESULT 28  
Q5B833 EMENI PRELIMINARY; PRT; 409 AA.  
AC Q5B833\_2005, integrated into UniProtKB/TrEMBL.  
DT 26-APR-2005, sequence version 1.  
DT 07-MAR-2006, entry version 6.  
DE Hypothetical protein.  
GN ORFNames=AN3297.2;  
OS Aspergillus nidulans FGSC A4.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; Emericella.  
OX NCBI\_TaxID=227321;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=FGSC 4;  
RX PubMed=16372000; DOI=10.1038/nature04341;  
RA Galagan J.E., Calvo S.E., Cuomo C., Ma L.-J., Wortman J.R., Batzoglou S., Lee S.-I., Basturkmen M., Spevak C.C., Clutterbuck J., Kapitonov V., Jurka J., Scanzocchio C., Farman M., Butler J., Purcell S., Harris S., Braus G.H., Draht O., Busch S., D'Enfert C., Boucher C., Goldman G.H., Bell-Pedersen D., Griffiths-Jones S., Doonan J.H., Yu J., Vienken K., Pain A., Freitag M., Selker E.U., Archer D.B., Penalva M.A., Oakley B.R., Momany M., Tanaka T., Kumagai T., Asai K., Machida M., Nierman W.C., Denning D.W., Caddick M., Hynes M., Paoletti M., Fischer R., Miller B.L., Dyer P.S., Sachs M.S., Omani S.A., Birren B.W.;  
RA "Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae."  
RL Nature 438:1105-1115(2005).  
CC  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
-----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
-----  
DR EMBL; AAC001000035; EAA63285.1; -; Genomic DNA.  
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR001547; Glyco\_hydro\_5.  
DR Pfam; PF00150; Cellulase; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 409 AA; 46168 MW; 93AB6EC788222720 CRC64;  
  
Query Match 54.9%; Score 45; DB 2; Length 409;  
Best Local Similarity 87.5%; Pred. No. 61;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 SWELGNBP 8  
Db 220 AWELGNBP 227  
:|||||  
-----  
RESULT 29  
Q99036 TRIRE PRELIMINARY; PRT; 437 AA.  
ID Q99036\_TRIRE PRELIMINARY; PRT; 437 AA.  
AC Q99036;  
DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.  
DT 01-NOV-1996, sequence version 1.  
DT 07-FEB-2006, entry version 26.  
DE Beta-mannase precursor.  
OS Trichoderma reesei (Hypocrea jecorina).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
OX NCBI\_TaxID=51453;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=RutC30;  
RA Stalbrand H., Saloheimo A., Vehmaanpera J., Penttila M.;  
RT "cDNA encoding Trichoderma reesei beta-mannanase."  
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
-----  
DR EMBL; L25310; AAA34208.1; -; mRNA.  
DR PDB; 1QNP; X-ray; A=28-371.  
DR PDB; 1QNP; X-ray; A=28-371.  
DR PDB; 1QNP; X-ray; A=28-371.  
DR PDB; 1QNP; X-ray; A=28-371.  
DR PDB; 1QNP; X-ray; A=28-371.  
DR PDB; 1QNP; X-ray; A=28-371.  
DR GO; GO:0005576; C:extracellular region; IEA.  
DR GO; GO:0030248; F:cellulose binding; IEA.  
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR000254; CBD\_fun.  
DR Pfam; PF00734; CBM\_1; 1.  
DR Pfam; PF00150; Cellulase; 1.  
DR ProDom; PD001821; CBD\_fungal; 1.  
DR SMART; SM00236; fCBD; 1.  
DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
KW Signal.  
FT SIGNAL 1 19 Potential.  
FT CHAIN 28 437 beta-mannase.  
SQ SEQUENCE 437 AA; 47053 MW; 17513DADE12654A7 CRC64;  
  
Query Match 54.9%; Score 45; DB 2; Length 437;  
Best Local Similarity 87.5%; Pred. No. 65;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 SWELGNBP 8  
Db 190 AWELGNBP 197  
:|||||  
-----  
RESULT 30  
Q3DZN2 CHLAU PRELIMINARY; PRT; 490 AA.  
ID Q3DZN2\_CHLAU PRELIMINARY; PRT; 490 AA.  
AC Q3DZN2;  
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.  
DT 22-NOV-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Hypothetical protein.  
GN ORFNames=CaurDRAFT\_3045;  
OS Chloroflexus aurantiacus J-10-fl.  
OC Bacteria; Chloroflexi; Chloroflexales; Chloroflexaceae; Chloroflexus.  
OX NCBI\_TaxID=324602;  
RN [1]

```

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=J-10-fl;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Chloroflexus
RT aurantiacus J-10-fl.";
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
[2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=J-10-fl;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Chloroflexus aurantiacus
RT J-10-fl.";
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
[3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=J-10-fl;
RG US DOE Joint Genome Institute;
RA Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
[4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=J-10-fl;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAH02000021; EA058861.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 490 AA; 53840 MW; 103222A5536FE211 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 490;
Best Local Similarity 87.5%; Pred. No. 74;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 WELGNEP 9
Db 163 WELGNEPD 170
|||||
|

RESULT 31
Q5CW71 CRYPV PRELIMINARY; PRT; 538 AA.
AC Q5CW71;
DT 12-APR-2005, integrated into UniProtKB/TrEMBL.
DT 12-APR-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE DNA primase large subunit (Fragment).
GN ORFNames:cgd8_1410;
OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=5807;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Iowa type II;
RX PubMed=15044751; DOI=10.1126/science.1094786;
RA Abrahamson M.S., Templeton T.J., Enomoto S., Abrahante J.E., Zhu G.,
RA Lantto C.A., Deng M., Liu C., Widmer G., Tzipori S., Buck G.A., Xu P.,
RA Bankier A.T., Dear P.H., Konfortov B.A., Spriggs H.F., Iyer L.,
RA Anantharaman V., Aravind L., Kapur V.;
RT "Complete genome sequence of the apicomplexan, Cryptosporidium
RT parvum.";
RL Science 304:441-445(2004).
CC -!- CAUTION: The sequence shown here is derived from an

```

```

CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAE01000003; EAK9678.1; -; Genomic_DNA.
DR GO; GO:0005658; C:alpha DNA polymerase:primase complex; IEA.
DR GO; GO:0003896; F:DNA primase activity; IEA.
DR GO; GO:0006269; P:DNA replication, synthesis of RNA primer; IEA.
DR InterPro; IPR007238; DNA_primase_lrg.
DR Pfam; PF04104; DNA_primase_lrg; I.
FT NON_TER 1
SQ SEQUENCE 538 AA; 63237 MW; 9AE256D7E36927C6 CRC64;

Query Match 54.9%; Score 45; DB 2; Length 538;
Best Local Similarity 58.3%; Pred. No. 82;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 LGNEPNSFLKKA 15
Db 514 LGNEPNSYFKKS 525
|||||
|

RESULT 32
BGLR MOUSE
ID BGLR MOUSE STANDARD; PRT; 648 AA.
AC P12265; Q61601; Q64473; Q64474.
DT 01-OCT-1989, integrated into UniProtKB/Swiss-Prot.
DT 01-OCT-1989, sequence version 1.
DT 21-FEB-2006, entry version 59.
DE Beta-glucuronidase precursor (EC 3.2.1.31).
GN Name=Gusb; Synonyms=Gus, Gus-s;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=88085188; PubMed=2891607;
RA Gallagher P.M., D'Amore M.A., Lund S.D., Elliott R.W., Pazik J.,
RA Hohman C., Korfhagen T.R., Ganschow R.E.;
RT "DNA sequence variation within the beta-glucuronidase gene complex
RT among inbred strains of mice.";
RL Genomics 1:145-152(1987).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=88284700; PubMed=3397060;
RA Gallagher P.M., D'Amore M.A., Lund S.D., Ganschow R.E.;
RT "The complete nucleotide sequence of murine beta-glucuronidase mRNA
RT and its deduced polypeptide.";
RL Genomics 2:215-219(1988).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=89062453; PubMed=3196706;
RA D'Amore M.A., Gallagher P.M., Korfhagen T.R., Ganschow R.E.;
RT "Complete sequence and organization of the murine beta-glucuronidase
RT gene.";
RL Biochemistry 27:7131-7140(1988).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C3H/HeJ and YBR; TISSUE=Sperm;
RX MEDLINE=89384641; PubMed=2779578;
RA Wawzyniak C.J., Gallagher P.M., D'Amore M.A., Carter J.E., Lund S.D.,
RA Rinchik E.M., Ganschow R.E.;
RT "DNA determinants of structural and regulatory variation within the
RT murine beta-glucuronidase gene complex.";
RL Mol. Cell. Biol. 9:4074-4078(1989).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=88216590; PubMed=2835664;
RA Funkenstein B., Leary S.L., Stein J.C., Catterall J.F.;

```

RT "Genomic organization and sequence of the Gus-s alpha allele of the  
 RL murine beta-glucuronidase gene.";  
 RN Mol. Cell. Biol. 8:1160-1168(1988).  
 RP NUCLEOTIDE SEQUENCE OF 593-648, VARIANT ARG-642, AND SUBCELLULAR  
 RP LOCATION.  
 RX MEDLINE=90368633; PubMed=2394691;  
 RA Li H., Takeuchi K.H., Manly K., Chapman V., Swank R.T.;  
 RA "The propogite of beta-glucuronidase. Further evidence of its  
 RT involvement in compartmentalization of beta-glucuronidase and sequence  
 RT similarity with portions of the reactive site region of the serpin  
 RT superfamily.";  
 RL J. Biol. Chem. 265:14732-14735(1990).  
 CC -!- FUNCTION: Plays an important role in the degradation of dermatan  
 CC and keratan sulfates.  
 CC -!- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = D-glucuronate  
 CC + an alcohol.  
 CC -!- SUBUNIT: Homotetramer.  
 CC -!- SUBCELLULAR LOCATION: Lysosomal. A small proportion is found in  
 CC the endoplasmic reticulum.  
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 2 family.  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 -----  
 DR EMBL: J03047; AAA37696.1; -; mRNA.  
 DR EMBL: J02836; AAA98623.1; -; Genomic\_DNA.  
 DR EMBL: M63836; AAA63309.1; -; mRNA.  
 DR EMBL: M28540; AAA63307.1; -; mRNA.  
 DR EMBL: M28541; AAA63308.1; -; mRNA.  
 DR EMBL: M19279; AAA37697.1; -; mRNA.  
 DR PIR: A32576; A32576.  
 DR HSP: P08236; 1BHG.  
 DR SRS: P12265; 22-628.  
 DR Ensembl: ENSMUSG00000025534; Mus musculus.  
 DR MGI: MGI:95872; Gusb.  
 DR InterPro: IPR006101; Glyco\_hydro\_2.  
 DR InterPro: IPR006102; Glyco\_hydro\_2lg.  
 DR InterPro: IPR006103; Glyco\_hydro\_2TIM.  
 DR InterPro: IPR006104; Glyco\_hydro\_s\_bd.  
 DR Pfam: PF00703; Glyco\_hydro\_2; 1.  
 DR Pfam: PF02836; Glyco\_hydro\_2\_C; 1.  
 DR Pfam: PF02837; Glyco\_hydro\_2\_N; 1.  
 DR PRINTS: PR00132; GLHYRIASE2.  
 DR PROSITE: PS00719; GLYCOSYL HYDROL\_F2\_1; 1.  
 DR PROSITE: PS00608; GLYCOSYL HYDROL\_F2\_2; 1.  
 DR KEGG: K01101; GLUCURONIDASE.  
 KW Endoplasmic reticulum; Glycoprotein; Glycosidase; Hydrolase; Lysosome;  
 KW Polymorphism; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 648  
 FT ACT SITE 447 447  
 FT CARBOHYD 172 172  
 FT CARBOHYD 416 416  
 FT CARBOHYD 591 591  
 FT CARBOHYD 627 627  
 FT VARIANT 87 87  
 FT VARIANT 233 233  
 FT VARIANT 265 265  
 FT VARIANT 320 320  
 FT VARIANT 428 428  
 FT VARIANT 616 616  
 FT VARIANT 642 642  
 FT SEQUENCE 648 AA; 74239 MW; 3B8C65A50B3B96D6 CRC64;  
 /FTID=PRO\_0000012162.  
 FT Proton donor (By similarity).  
 FT N-linked (GlcNAc...) (Potential).  
 FT N-linked (GlcNAc...) (Potential).  
 FT N-linked (GlcNAc...) (Potential).  
 FT N-linked (GlcNAc...) (Potential).  
 FT I -> T (in strain: C3H/HeJ).  
 FT I -> T (in allele GUS-SA).  
 FT D -> G (in strain: YBR and C3H/HeJ).  
 FT V -> I (in strain: YBR and C3H/HeJ).  
 FT E -> K (in allele GUS-SA).  
 FT F -> L (in allele GUS-SA).  
 FT G -> R (in allele W26; reduced retention  
 in the endoplasmic reticulum).  
 Query Match 54.9%; Score 45; DB 1; Length 648;  
 Best Local Similarity 57.1%; Pred. NO. 1e+02;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Qy 2 WELNEPNSFLKKA 15  
 | : ||| : |||  
 Db 442 WSVANEPSALKKA 455

RESULT 33  
 Q3TW82\_MOUSE  
 ID Q3TW82\_MOUSE PRELIMINARY; PRT; 648 AA.  
 AC Q3TW82;  
 DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.  
 DT 07-FEB-2006, entry version 5.  
 DE Osteoclast-like cell cDNA, RIKEN full-length enriched library,  
 DE clone:I420031K20 product:beta-glucuronidase, full insert sequence  
 DE (Osteoclast-like cell cDNA, RIKEN full-length enriched library,  
 DE clone:I420006F06 product:beta-glucuronidase, full insert sequence).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Methods Enzymol. 303:19-44(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Bremner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
 RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bersano T., Bono H., Chalk A.M.,  
 RA Bansal K.P., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Hill D., Hummel L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
 RA Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 RA Mottagui-Tabar S., Mulder N., Nakano N., Nakaguchi H., Ng P.,  
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavoni G., Pesole G.,  
 RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,  
 RA Tamajima K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 RA Fukuda S., Kanamori-Kiyama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
 RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome.";  
 RL Science 303:1559-1563(2005).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RG RIKEN Genome Exploration Research Group, and Genome Science Group

RG (Genome Network Core Team) and the PANTOM Consortium;  
 RT "Antisense Transcription in the Mammalian Transcriptome.";  
 RN Science 309:1564-1566(2005).  
 RP [4]

# RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nakai I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schraml L.M., Kanpin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Giasi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Perlee G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirose-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs";  
 RL Nature 420:563-573(2002).  
 RN [5]

# RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schraml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RT Nature 409:685-690(2001).  
 RN [6]

# RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [7]

# RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;

RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Suno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,  
 RA Kuni N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [8]

# RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J;  
 RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
 RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
 RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
 RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
 RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
 RA Muramatsu M., Hayashizaki Y.;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = D-glucuronate  
 CC + an alcohol.

# CC -!- SUBUNIT: Homotrimer (By similarity).

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonCommercial License  
 CC -----  
 CC EMBL; AK159804; BAE35384.1; -; mRNA.  
 DR EMBL; AK159163; BAE34865.1; -; mRNA.  
 DR GO; GO:0005764; C:lysosome; RCA.  
 DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; RCA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; RCA.  
 DR InterPro; IPR006101; Glyco\_hydro\_2.  
 DR InterPro; IPR006102; Glyco\_hydro\_21g.  
 DR InterPro; IPR006103; Glyco\_hydro\_2T1M.  
 DR InterPro; IPR006104; Glyco\_hydro\_5\_bd.  
 DR Pfam; PF00703; Glyco\_hydro\_2; 1.  
 DR Pfam; PF02836; Glyco\_hydro\_2\_C; 1.  
 DR Pfam; PF02837; Glyco\_hydro\_2\_N; 1.  
 DR PRINTS; PR00132; GLHYDRLASE2.  
 DR PROSITE; PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
 DR PROSITE; PS00609; GLYCOSYL\_HYDROL\_F2\_2; 1.  
 DR Lysosome.

SQ SEQUENCE 648 AA; 74271 MW; 93F59C49BB186BB5 CRC64;

Query Match Similarity 54.9%; Score 45; DB 2; Length 648;

Best Local Similarity 57.1%; Pred. No. le+02; Mismatches 4; Indels 0; Gaps 0;

Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 WELGNEPNSFLKKA 15

Db 442 WSVANEPSSALKPA 455

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Q6IR10 MOUSE PRELIMINARY; PRT; 648 AA.

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muroidea; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA STRAIN=NMRI; TISSUE=Mammary tumor. WAP-Tag model. 5 months old;  
 RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shmolen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heisch F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Brownstein M., Soares M.B., Donald M.P., Casavant T.L., Scheetz T.E.,  
 RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwen P.J., McKernan K.J., Carninci P., Prange C.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Griinwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smal M.A.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RL and mouse cDNA sequences.";  
 RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RA STRAIN=NMRI; TISSUE=Mammary tumor. WAP-Tag model. 5 months old;  
 RC Director MGC Project;  
 RA Submitted (JUN-2004) to the ENBL/GenBank/DBJ databases.  
 [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RA STRAIN=C57BL/6J; TISSUE=Bone marrow, Cecum, and Embryonic body between  
 RC diaphragm region and neck;  
 RX PubMed=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Methods Enzymol. 303:19-44 (1999).  
 [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RA STRAIN=C57BL/6J; TISSUE=Bone marrow, Cecum, and Embryonic body between  
 RC diaphragm region and neck;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Lenhard T., Welle C., Kodzius R., Shimokawa K.,  
 RA Oyama R., Ravasi T., Lenhard T., Batalov S., Renshaw A., Bailey T.L.,  
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 RA di Bernardo D., Down T., Engstrom P., Fgiolini M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA Georgi-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Hill D., Huminecki L., Iacono M., Ieko K., Imawa A., Ishikawa T.,  
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
 RA Kitano H., Kollas L., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Larau L.F., Lazarevic D., Lipovich L., Liu J.,  
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 RA Motagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,  
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
 RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
 RA Roost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sesca L., Sheng Y.,  
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,  
 RA Tamofa K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,

Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kishimoto T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
 RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome.";  
 RL Science 309:1559-1563 (2005).  
 [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RA STRAIN=C57BL/6J; TISSUE=Bone marrow, Cecum, and Embryonic body between  
 RC diaphragm region and neck;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RA RIKEN Genome Exploration Research Group, and Genome Science Group  
 (Genome Network Core Team) and the FANTOM Consortium;  
 RG "Antisense Transcription in the Mammalian Transcriptome.";  
 RL Science 309:1564-1566 (2005).  
 [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RA STRAIN=C57BL/6J; TISSUE=Bone marrow, Cecum, and Embryonic body between  
 RC diaphragm region and neck;  
 RX MEDLINE=22354663; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer J.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Inotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RL 60,770 full-length cDNAs.";  
 RX Nature 420:563-573 (2002).  
 [7]  
 RP NUCLEOTIDE SEQUENCE.  
 RA STRAIN=C57BL/6J; TISSUE=Bone marrow, Cecum, and Embryonic body between  
 RC diaphragm region and neck;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bargh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Bone marrow, Cecum, and Embryonic body between  
RC diaphragm region and neck;  
RC MEDLINE=20493374; PubMed=11042159; DOI=10.1101/gr.145100;  
RX Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu N., Hayashizaki Y.,  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [9]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Bone marrow, Cecum, and Embryonic body between  
RC diaphragm region and neck;  
RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [10]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Bone marrow, and Cecum;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,

Query Match 54.9%; Score 45; DB 2; Length 648;  
Best Local Similarity 57.1%; Pred. No. 1e+02;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
DB 442 WSVANEPSALKPA 455

QY 2 WELGNPEPSFLKKA 15  
| : ||| : |||  
DB 442 WSVANEPSALKPA 455

RESULT 35  
Q4IR74 GIBZE PRELIMINARY; PRT; 1687 AA.  
AC Q4IR74;  
DT 16-AUG-2005, integrated into UniProtKB/TrEMBL.  
DT 16-AUG-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Hypothetical protein.  
OS ORFNames=FG00284.1;  
GN Gibberella zeae (Fusarium graminearum).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
OX NCBI\_TaxID=5518;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=PH-1 / NRRL 31084;  
RA Birren B.W., Nusbaum C., Abouelleil A., Allen N., Anderson S.,  
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,  
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,  
RA Choel Y., Collymore A., Cook A., Cooke P., Corum B., DeArelano K.,  
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,  
RA Erickson J., Faro S., Ferreira P., Fitzgerald M., Gage D.,  
RA Galagan J.E., Gardyna S., Gnerre S., Graham L., Grand-Pierre N.,  
RA Hafez N., Hagopian D., Hagos B., Hall J., Horton L., Hulme W.,  
RA Iliev I., Jaffe D., Johnson R., Jones C., Kamal M., Kamat A.,  
RA Karatas A., Kellis C., Landers T., Levine R., Lindblad-Toh K., Liu G.,

RA Lui A., Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J.,  
RA Manning J., Matthews C., Mauceli E., McCarthy M., Meldrim J.,  
RA Meneus L., Mihova T., Mienga V., Murphy T., Naylor J., Nguyen C.,  
RA Nicol R., Nielsen C.B., Norbu C., O'Connor T., O'Donnell P.,  
RA O'Neill D., Oliver J., Peterson K., Phunkhang P., Pierre N.,  
RA Purcell S., Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C.,  
RA Rogov P., Roman J., Schauer S., Schupack R., Seaman S., Severy P.,  
RA Smirnov S., Smith C., Spencer B., Stange-Thomann N., Stojanovic N.,  
RA Stubbs M., Talmas J., Testaye S., Theodore J., Topham K., Travers M.,  
RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,  
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,  
RA Lander E.S.;  
RT "Fusarium graminearum genome sequence.";  
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL: AACM01000010; EAA69694.1; -: Genomic\_DNA.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 1687 AA; 189392 MW; 9DFAL3423199466C CRC64;  
Query Match 54.9%; Score 45; DB 2; Length 1687;  
Best Local Similarity 50.0%; Pred. No. 2.9e+02;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 1 SWELGNPEPSFLKK 14  
| : ||| : |||  
DB 1106 SWILGNBPGVINE 1119

RESULT 36  
Q3D554 STRAG PRELIMINARY; PRT; 90 AA.  
AC Q3D554;  
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.  
DT 22-NOV-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Hypothetical protein.  
GN ORFNames=SAN 2117;  
OS Streptococcus agalactiae COH1.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=342616;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=COH1;  
RX PubMed=16172379; DOI=10.1073/pnas.0506758102;  
RA Tettelin H., Maignani V., Cieslewicz M.J., Donati C., Medini D.,  
RA Ward N.L., Angiuoli S.V., Crabtree J., Jones A.L., Durkin A.S.,  
RA DeBoy R.T., Davidson T.M., Mora M., Scarcelli M., Margarit y Ros I.,  
RA Peterson J.D., Hauser C.R., Sundaram J.P., Nelson W.C., Madupu R.,  
RA Brinkac L.M., Dodson R.J., Rosovitz M.J., Sullivan S.A.,  
RA Daugherty S.C., Haft D.H., Selengut J., Gwinn M.L., Zhou L., Zafar N.,  
RA Khouri H., Radune D., Dmitrov G., Watkins K., O'Connor K.J.,  
RA Smith S., Utterback T.R., White O., Rubens C.E., Grandi G.,  
RA Madoff L.C., Kasper D.L., Telford J.L., Wessels M.R., Rappuoli R.,  
RA Fraser C.M.;  
RT "Genome analysis of multiple pathogenic isolates of Streptococcus  
RT agalactiae: implications for the microbial 'pan-genome'";  
RL Proc. Natl. Acad. Sci. U.S.A. 102:13950-13955(2005).  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL: AAJ01000246; EAO74619.1; -: Genomic\_DNA.  
KW Hypothetical protein.



```
SQ SEQUENCE 90 AA; 10511 MW; 5E2CDE355801D9B4 CRC64;
Query Match 53.7%; Score 44; DB 2; Length 90;
Best Local Similarity 64.3%; Pred. No. 17;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SWELGNPNFLKK 14
   ||| ||| |||
Db 31 SWEKAGELNSLLKK 44

RESULT 37
Q3DNV8_STRAG PRELIMINARY; PRT; 90 AA.
AC Q3DNV8;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 22-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Hypothetical protein.
GN ORFNames=SAL_2048;
OS Streptococcus agalactiae 515.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=342614;
   [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=515;
RX PubMed=16172379; DOI=10.1073/pnas.0506758102;
RA Tettelin H., Maignani V., Cieslewicz M.J., Donati C., Medini D.,
RA Ward N.L., Angiuoli S.V., Crabtree J., Jones A.L., Durkin A.S.,
RA DeBoy R.T., DavidSEN T.M., Mora M., Scarselli M., Margarit Y Ros I.,
RA Peterson J.D., Hauser C.R., Sundaram J.P., Nelson W.C., Madupu R.,
RA Brinkac L.M., Dodson R.J., Rosovitz M.J., Sullivan S.A.,
RA Daugherty S.C., Haft D.H., Selengut J., Gwinn M.L., Zhou L., Zafar N.,
RA Khouri H., Radune D., Dimitrov G., Watkins K., O'Connor K.J.,
RA Smith S., Utterback T.R., White O., Rubens C.E., Grandi G.,
RA Madoff L.C., Kasper D.L., Telford J.L., Wessels M.R., Rappuoli R.,
RA Fraser C.M.;
RT "Genome analysis of multiple pathogenic isolates of Streptococcus
RT agalactiae: implications for the microbial 'pan-genome'.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:13950-13955(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDAJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AALP01000002; EAO72141.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 90 AA; 10568 MW; 57254EAC5198D9BB CRC64;
Query Match 53.7%; Score 44; DB 2; Length 90;
Best Local Similarity 64.3%; Pred. No. 17;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SWELGNPNFLKK 14
   ||| ||| |||
Db 31 SWEKAGELNSLLKK 44

RESULT 38
Q48S82_STRPM PRELIMINARY; PRT; 90 AA.
AC Q48S82;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE Hypothetical protein.
GN OrderedLocNames=M28_Spy1318;
OS Streptococcus pyogenes serotype M28.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
```

```
OX NCBI_TaxID=319700;
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=MQA86180 / Serotype M28;
RX PubMed=16088825; DOI=10.1086/430618;
RA Green N.M., Zhang S., Porcella S.F., Nagiec M.J., Barbian K.D.,
RA Beres S.B., Lefebvre R.B., Musser J.M.;
RT "Genome sequence of a serotype M28 strain of group A Streptococcus:
RT potential new insights into puerperal sepsis and bacterial disease
RT specificity.";
RL J. Infect. Dis. 192:760-770(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000056; AAX72428.1; -; Genomic_DNA.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 90 AA; 10568 MW; 5694EC2EC18B5A2B CRC64;
Query Match 53.7%; Score 44; DB 2; Length 90;
Best Local Similarity 64.3%; Pred. No. 17;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SWELGNPNFLKK 14
   ||| ||| |||
Db 31 SWEKAGELNSLLKK 44

RESULT 39
Q8DX45_STRAS PRELIMINARY; PRT; 90 AA.
AC Q8DX45;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 9.
DE Hypothetical protein SAG2012.
GN OrderedLocNames=SAG2012;
OS Streptococcus agalactiae serotype V.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
   [1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=2603 V/R / ATCC BAA-511 / Serotype V;
RX MEDLINE=22222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.N.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB014282; AAN00871.1; -; Genomic_DNA.
DR TIGR; SAG2012; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 90 AA; 10511 MW; 5E2CDE355801D9B4 CRC64;
Query Match 53.7%; Score 44; DB 2; Length 90;
Best Local Similarity 64.3%; Pred. No. 17;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 SWELGNPNFLKK 14
   ||| ||| |||
Db 31 SWEKAGELNSLLKK 44
```





```

CC  -!- SIMILARITY: Belongs to the nucleosome assembly protein (NAP)
CC  family.
CC  -----
CC  Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC  Distributed under the Creative Commons Attribution-NoDerivs License
CC  -----
DR  EMBL; AF042180; AAC62383.1; -; mRNA.
DR  EMBL; AK053812; BAC35536.1; -; mRNA.
DR  EMBL; BC011213; AAH11213.1; -; mRNA.
DR  Ensembl; ENSMUSG00000047514; Mus musculus.
DR  MGI; MGI:1298395; Tpy11.
DR  InterPro; IPR002164; NAP family.
DR  PANTHER; PTHR11875; NAP_family; 1.
DR  Pfam; PF00956; NAP; 1.
KW  Nuclear protein.
FT  CHAIN 1 379 Testis-specific Y-encoded-like protein 1.
FT  FTID=PRO_0000185671.
SQ  SEQUENCE 379 AA; 42994 MW; 92866C9CA557FC2D CRC64;

Query Match 53.7%; Score 44; DB 1; Length 379;
Best Local Similarity 46.2%; Pred. NO. 84;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 WELGNENSPFLKK 14
   | :|:| |:::
DB 299 WRRGHEPQSFIIR 311

RESULT 42
Q3TKXO_MOUSE
ID Q3TKXO_MOUSE PRELIMINARY; PRT; 379 AA.
AC Q3TKXO;
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 11-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE Blastocyst blastocyst cDNA, RIKEN full-length enriched library,
DE clone:11C0014007 product:testis-specific protein, Y-encoded-like, full
DE insert sequence (Blastocyst blastocyst cDNA, RIKEN full-length
DE enriched library, clone:11C0020N16 product:testis-specific protein, Y-
DE encoded-like, full insert sequence) (Osteoclast-like cell cDNA, RIKEN
DE full-length enriched library, clone:1420019020 product:testis-specific
DE protein, Y-encoded-like, full insert sequence).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
PY [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J;
RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
CA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Methods Enzymol. 303:19-44 (1999).
[2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J;
RC PubMed=16141072; DOI=10.1126/science.1112014;
CA Carninci P., Katukawa T., Katayama S., Gough J., Frich M.C., Maeda N.,
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
RA Ambesi-Impicciato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA Chiu K.P., Choudhary V., Christoffels A., Cluterbuck D.R.,
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
RA Hill D., Humintzki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
RA Jakt M., Kanapi A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,

```

Kurochkin I. V., Lareau L.F., Lazarevic D., Lipovich L., Liu J., Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L., Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K., Mottaqui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P., Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O., Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G., Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M., Roost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C., Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y., Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B., Sperling S., Stupka E., Sugiyura K., Sultana R., Takenaka Y., Taki K., Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A., Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K., Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C., Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J., Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y., Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T., Iida J., Inamura K., Itoh M., Kato T., Kawaji H., Kawagashira N., Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N., Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S., Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J., Hayashizaki Y.;

"The transcriptional landscape of the mammalian genome.";

Science 309:1559-1563 (2005).

[3]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J;

RX PubMed=163141073; DOI=10.1126/science.1112009;

RG RIKEN Genome Exploration Research Group, and Genome Science Group (Genome Network Core Team) and the FANTOM Consortium;

RT "Antisense Transcription in the Mammalian Transcriptome.";

RL Science 309:1564-1566 (2005).

[4]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J;

RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;

RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S., Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T., Baldairelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Bradt D., Brusic V., Chochia L.C., Corbani L.E., Cousins S., Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Fraser K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanaai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Meglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Nagashima T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M., Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M., Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L., Yuan Z., Zavalon M., Zhu Y., Zimmer A., Carninci P., Hayatsu N., Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K., Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A., Yamanishi A., Yoshino M., Wataston R., Lander E.S., Rogers J., Birney E., Hayashizaki Y.;

"Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";

Nature 420:563-573 (2002).

[5]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J;

RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;

RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schram L.M., Stauble F., Suzuki R., Tonita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Sult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,  
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RL "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsumai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.  
RN [9]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
RN [10]  
RP NUCLEOTIDE SEQUENCE.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AK166805; BAE39033.1; -; mRNA.  
CC EMBL; AK166893; BAE39098.1; -; mRNA.  
CC EMBL; AK159432; BAE35079.1; -; mRNA.  
CC GO; GO:0005634; C:nucleus; RCA.  
CC GO; GO:0003677; F:DNA binding; RCA.  
CC GO; GO:0006334; P:nucleosome assembly; RCA.  
CC InterPro; IPR002164; NAP\_family.  
CC Pfam; PF00956; NAP; 1.  
CC SEQUENCE 379 AA; 42994 MW; 9286C9CAA557FC2D CRC64;

Query Match 53.7%; Score 44; DB 2; Length 379;  
Best Local Similarity 46.2%; Pred. No. 84;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
QY 2 WELGNPNFLKK 14  
DB 299 WRRGEPQSFIRR 311  
RESULT 43  
Q642B1 RAT  
ID Q642B1 RAT PRELIMINARY; PRT; 379 AA.  
AC Q642B1;  
DT 25-OCT-2004, integrated into UniProtKB/TrEMBL.  
DT 07-FEB-2006, entry version 1.  
DE Testis-specific protein, Y-encoded-like (Predicted).  
GN Name=tepyl;  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridea; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Lung;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Tohiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RL and mouse cDNA sequences.";  
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Lung;  
RX NIH MGC Project;  
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.  
RN [3]  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; BC081955; AAH81955.1; -; mRNA.  
CC Ensembl; ENSRNOG0000000549; Rattus norvegicus.  
CC GO; GO:0005634; C:nucleus; IEA.  
CC GO; GO:0006334; P:nucleosome assembly; IEA.  
CC InterPro; IPR002164; NAP\_family.  
CC PANTHER; PTHR11875; NAP\_family; 1.  
CC Pfam; PF00956; NAP; 1.  
CC SEQUENCE 379 AA; 42704 MW; 08E0F0AD37FC1E4 CRC64;  
Query Match 53.7%; Score 44; DB 2; Length 379;  
Best Local Similarity 46.2%; Pred. No. 84;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
QY 2 WELGNPNFLKK 14  
DB 299 WRRGEPQSFIRR 311

```

RESULT 44
Q5TFE6 HUMAN PRELIMINARY; PRT; 437 AA.
ID Q5TFE6;
AC Q5TFE6;
DT 21-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 21-DEC-2004, sequence version 1.
DT 07-FEB-2006, entry version 7.
DE OTTHUMP00000017052.
GN Name=TSPL; ORFNames=RP3-486I3.3-001;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Barlow K.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AL050331; CAB55883.1; -; Genomic DNA.
DR Ensembl; ENSG00000189241; Homo sapiens.
DR LinkHub; Q5TFE6; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0006334; P:nucleosome assembly; IEA.
DR InterPro; IPR002164; NAP family.
DR PANTHER; PTHR11875; NAP_Family; 1.
DR Pfam; PF00956; NAP; 1.
SQ SEQUENCE 437 AA; 49192 MW; D651AGFAA7CF811F CRC64;

Query Match 53.7%; Score 44; DB 2; Length 437;
Best Local Similarity 46.2%; Pred. No. 98;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 2 WELGNERNFLKK 14
| : : : : :
Db 357 WRRGHEPQSFIIR 369

RESULT 45
TSYL1 HUMAN STANDARD; PRT; 438 AA.
ID TSYL1 HUMAN STANDARD; PRT; 438 AA.
AC QH0U9; O75885;
DT 15-AUG-2003, integrated into UniProtKB/Swiss-Prot.
DT 15-AUG-2003, sequence version 2.
DE Testis-specific Y-encoded-like protein 1 (TSPLY-like 1).
DE Name=TSYL1; Synonyms=TSPLY;
GN Name=TSYL1; Synonyms=TSPLY;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Fetal brain;
RX MEDLINE=21154917; PubMed=11230166; DOI=10.1101/gr.154701;
RX Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Anorge W., Boecker M., Bloecker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,
RA Mews H.-W., Ottenwaelter B., Oberwälder B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RT "Towards a catalog of human genes and proteins: sequencing and
RT analysis of 500 novel complete protein coding human cDNAs.";
RL Genome Res. 11:422-435(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

```

```

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE [MRNA] OF 366-438, AND TISSUE SPECIFICITY.
RX MEDLINE=98399864; PubMed=9730615;
RA Vogel T., Dittich O., Mehraein Y., Dechend F., Schnieders F.,
RA Schmidtke J.;
RT "Murine and human TSPYL genes: novel members of the TSPY-SET-NAP1L1
RT family.";
RL Cytogenet. Cell Genet. 81:265-270(1998).
RN [4]
RP SUBCELLULAR LOCATION, AND IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=22317277; PubMed=12429849; DOI=10.1091/mbc.E02-05-0271;
RA Scherl A., Coute Y., Deon C., Calle A., Kindbeiter K., Sanchez J.-C.,
RA Greco A., Hochstrasser D.F., Diaz J.-J.;
RT "Functional proteomic analysis of human nucleolus.";
RL Mol. Biol. Cell 13:4100-4109(2002).
RN [5]
RP INVOLVEMENT IN SIDDT.
RX PubMed=15273283; DOI=10.1073/pnas.0401194101;
RA Puffenberger E.G., Hu-Lince D., Parod J.M., Craig D.W., Dobrin S.E.,
RA Conway A.R., Donarum E.A., Strauss K.A., Dunkley T., Cardenas J.F.,
RA Melmed K.R., Wright C.A., Liang W., Stafford P., Flynn C.R.,
RA Morton D.H., Stephan D.A.;
RT "Mapping of sudden infant death with dysgenesis of the testes syndrome
RT (SIDDT) by a SNP genome scan and identification of TSPYL loss of
RT function.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:11689-11694(2004).
RN [6]
RP SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -!- TISSUE SPECIFICITY: Expressed in testis, ovary, liver, spleen,
CC -!- brain, kidney, prostate, lung, liver, and heart.
CC -!- DISEASE: Defects in TSPYL are the cause of sudden infant death
CC with dysgenesis of the testes syndrome (SIDDT) [MIM:608800]. SIDDT
CC is an autosomal recessive disorder. Affected infants appear normal
CC at birth, develop signs of viscerotautonomic dysfunction early in
CC life, and die before 12 months of age of abrupt cardiorespiratory
CC arrest. Features included bradycardia, hypothermia, severe
CC gastroesophageal reflux, laryngospasm, bronchospasm, and abnormal
CC cardiorespiratory patterns during sleep. Genotypic males with
CC SIDDT had fetal testicular dysgenesis and ambiguous genitalia,
CC with findings such as intraabdominal testes, dysplastic testes,
CC deficient fetal testosterone production, fusion and rugation of
CC the gonadal sac, and partial development of the penile shaft.
CC Female sexual development was normal. Affected infants had an
CC unusual staccato cry, similar to the cry of a goat.
CC -!- SIMILARITY: Belongs to the nucleosome assembly protein (NAP)
CC family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AL136629; CAB66564.1; -; mRNA.
DR EMBL; BC048969; AAH48969.1; -; mRNA.
DR EMBL; AF042181; AAC62384.1; -; mRNA.
DR SWISS-2DPAGE; Q9H0U9; HUMAN.
DR Ensembl; ENSG00000189241; Homo sapiens.

```

```
DR HGNC; HGNC:12382; TSPYL1.
DR MIM; 604714; gene.
DR MIM; 608800; phenotype.
DR InterPro; IPR002164; NAP_family.
DR PANTHER; PTHR11875; NAP_family; 1.
DR Pfam; PF00956; NAP; 1.
KW Nuclear protein; Polymorphism.
FT CHAIN 1 438
FT VARIANT 62 62 P -> S (in dbSNP:3828743).
FT VARIANT 74 74 A -> P (in dbSNP:3749895).
FT SEQUENCE 438 AA; 49292 MW; 238B9FB17736116D CRC64;
Query Match 53.7%; Score 44; DB 1; Length 438;
Best Local Similarity 46.2%; Pred. No. 98;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 2 WELGNPNFLKK 14
Db 358 WRRGHEPQSFIIR 370
RESULT 46
Q6FI91_HUMAN PRELIMINARY; PRT; 438 AA.
AC Q6FI91;
DT 19-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE TSPYL protein.
GN Name=TSPYL;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Ebert L., Schick M., Neubert P., Schatten R., Henze S., Korn B.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CR533535; CAG38566.1; -; mRNA.
DR Ensembl; ENSG00000189241; Homo sapiens.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0006334; P:nucleosome assembly; IEA.
DR InterPro; IPR002164; NAP_family.
DR PANTHER; PTHR11875; NAP_family; 1.
DR Pfam; PF00956; NAP; 1.
SQ SEQUENCE 438 AA; 49307 MW; 834D4B7B72B59650 CRC64;
Query Match 53.7%; Score 44; DB 2; Length 438;
Best Local Similarity 46.2%; Pred. No. 98;
Matches 6; Conservative 4; Mismatches 4; Indels 3; Indels 0; Gaps 0;
QY 2 WELGNPNFLKK 14
Db 358 WRRGHEPQSFIIR 370
RESULT 47
Q5R5G8_PONPY PRELIMINARY; PRT; 438 AA.
AC Q5R5G8;
DT 21-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 21-DEC-2004, sequence version 1.
DT 07-FEB-2006, entry version 6.
DE Hypothetical protein DKFP469F1920.
GN Name=DKFP469F1920;
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Bloeker H., Boecher M., Brandt P., Mewes H.W., Weil B., Amid C.,
RA Osanger A., Fobo G., Han M., Wiemann S.;
RL Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CR859130; CAH91322.1; -; mRNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0006334; P:nucleosome assembly; IEA.
DR InterPro; IPR002164; NAP_family.
DR PANTHER; PTHR11875; NAP_family; 1.
DR Pfam; PF00956; NAP; 1.
KW Hypothetical protein.
SQ SEQUENCE 438 AA; 49350 MW; 834D4B7B72B59650 CRC64;
Query Match 53.7%; Score 44; DB 2; Length 438;
Best Local Similarity 46.2%; Pred. No. 98;
Matches 6; Conservative 4; Mismatches 4; Indels 3; Indels 0; Gaps 0;
QY 2 WELGNPNFLKK 14
Db 358 WRRGHEPQSFIIR 370
RESULT 48
Q5RA88_PONPY PRELIMINARY; PRT; 438 AA.
AC Q5RA88;
DT 21-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 21-DEC-2004, sequence version 1.
DT 07-FEB-2006, entry version 6.
DE Hypothetical protein DKFP459D1728.
GN Name=DKFP459D1728;
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Bloeker H., Boecher M., Brandt P., Mewes H.W., Weil B., Amid C.,
RA Osanger A., Fobo G., Han M., Wiemann S.;
RL Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CR859130; CAH91322.1; -; mRNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0006334; P:nucleosome assembly; IEA.
DR InterPro; IPR002164; NAP_family.
DR PANTHER; PTHR11875; NAP_family; 1.
DR Pfam; PF00956; NAP; 1.
KW Hypothetical protein.
SQ SEQUENCE 438 AA; 49350 MW; 834D4B7B72B59650 CRC64;
Query Match 53.7%; Score 44; DB 2; Length 438;
Best Local Similarity 46.2%; Pred. No. 98;
Matches 6; Conservative 4; Mismatches 4; Indels 3; Indels 0; Gaps 0;
QY 2 WELGNPNFLKK 14
Db 358 WRRGHEPQSFIIR 370
```

```
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Bloeker H., Boecher M., Brandt P., Mewes H.W., Weil B., Amid C.,
RA Osanger A., Fobo G., Han M., Wiemann S.;
RL Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CR859130; CAH92998.1; -; mRNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0006334; P:nucleosome assembly; IEA.
DR InterPro; IPR002164; NAP_family.
DR PANTHER; PTHR11875; NAP_family; 1.
DR Pfam; PF00956; NAP; 1.
KW Hypothetical protein.
SQ SEQUENCE 438 AA; 49292 MW; F885B1D3D455F39B CRC64;
Query Match 53.7%; Score 44; DB 2; Length 438;
Best Local Similarity 46.2%; Pred. No. 98;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 2 WELGNPNFLKK 14
Db 358 WRRGHEPQSFIIR 370
RESULT 49
Q5RA88_PONPY PRELIMINARY; PRT; 438 AA.
AC Q5RA88;
DT 21-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 21-DEC-2004, sequence version 1.
DT 07-FEB-2006, entry version 6.
DE Hypothetical protein DKFP459D1728.
GN Name=DKFP459D1728;
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Bloeker H., Boecher M., Brandt P., Mewes H.W., Weil B., Amid C.,
RA Osanger A., Fobo G., Han M., Wiemann S.;
RL Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CR859130; CAH91322.1; -; mRNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0006334; P:nucleosome assembly; IEA.
DR InterPro; IPR002164; NAP_family.
DR PANTHER; PTHR11875; NAP_family; 1.
DR Pfam; PF00956; NAP; 1.
KW Hypothetical protein.
SQ SEQUENCE 438 AA; 49350 MW; D4B74PFC42D9E066 CRC64;
Query Match 53.7%; Score 44; DB 2; Length 438;
Best Local Similarity 46.2%; Pred. No. 98;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 2 WELGNPNFLKK 14
Db 358 WRRGHEPQSFIIR 370
```

```

Db          | :||| ||:::
358 WRGHEPQSFIRR 370

RESULT 49
QSQNQ4 MOUSE
ID QSQNQ4 MOUSE PRELIMINARY; PRT; 497 AA.
AC QSQNQ4;
DT 04-JAN-2005, integrated into UniProtKB/TrEMBL.
DT 04-JAN-2005, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Oxyeterol-binding protein.
DE Oxyeterol-binding protein.
GN Name=Osbp2; ORFNames=RP23-309E11.7-003;
OS Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Bird C.;
RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
[2]
RP NUCLEOTIDE SEQUENCE.
RA Dunn M.;
RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AL691413; CAI25059.1; -; Genomic_DNA.
DR EMBL; AL731853; CAI25059.1; JOINED; Genomic_DNA.
DR EMBL; AL731853; CAI51855.1; -; Genomic_DNA.
DR EMBL; AL691413; CAI51855.1; JOINED; Genomic_DNA.
DR Ensembl; ENSMUSG0000020435; Mus musculus.
DR GO; GO:0006869; P:lipid transport; IEA.
DR GO; GO:0008202; P:steroid metabolism; IEA.
DR InterPro; IPR000648; Oxysterol bd.
DR PANTHER; PTHR10972; Oxysterol BP; 1.
DR Pfam; PF01237; Oxysterol BP; 1.
DR PROSITE; PS01013; OSBP; 1.
DR Lipid transport; Transport.
KW SEQUENCE 497 AA; 57076 MW; 1BE492F0B494223D CRC64;
SQ
Query Match 53.7%; Score 44; DB 2; Length 497;
Best Local Similarity 72.7%; Pred. No. 1.1e+02;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 WELGNPNPFL 12
||| :|||
Db 14 WELGWKPGSFL 24

RESULT 50
Q47BW1 DECAR
ID Q47BW1 DECAR PRELIMINARY; PRT; 492 AA.
AC Q47BW1;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Conserved hypothetical transmembrane protein.
OS OrderedLocusNames=Daro_2940;
GN Oxyeterol-binding protein.
OC Bacteria; Proteobacteria; Betaproteobacteria; Rhodocyclales;
OC Rhodocyclaceae; Dechloromonas.
OX NCBI_TaxID=159087;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RA Copeland A., Lucas S., Lapidus A., Barty K., Better J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Di Bartolo G., Trono K., Kellar K.,
RA Schmutz J., Larimer F., Land M., Ivanova N., Richardson P.;
RL "Complete sequence of Dechloromonas aromatica RCB.;"
RL submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.

```

GenCore version 5.1.1-9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:31:47 ; Search time 132.219 Seconds

(without alignments)

65.702 Million cell updates/sec

Title: US-10-645-659A-7

Perfect score: 104

Sequence: 1 PAYLRFQGTXTDFLIFDPK 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

A\_Geneseq\_8:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

9: Geneseq2005s:\*

10: Geneseq2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	100.0	19	8	ADR88213
2	104	100.0	19	8	ADT78180
3	104	100.0	19	8	AEA42429
4	104	100.0	74	8	ADR88217
5	104	100.0	74	8	ADT78184
6	104	100.0	74	8	ADY27059
7	104	100.0	74	8	ADZ18994
8	104	100.0	74	8	AEA42433
9	104	100.0	174	9	ADZ19012
10	104	100.0	460	9	ADY27061
11	104	100.0	486	9	ABE87589
12	104	100.0	492	9	ADZ18996
13	104	100.0	493	9	ABE87562
14	104	100.0	495	9	ADZ18999
15	104	100.0	497	9	ABE87587
16	104	100.0	501	9	ADZ19000
17	104	100.0	507	9	ADZ19005
18	104	100.0	508	9	ADY27058
19	104	100.0	526	9	ADZ19006
20	104	100.0	527	5	ABE07815
21	104	100.0	527	7	ABW02018
22	104	100.0	527	8	ADO63825
23	104	100.0	527	8	ADO63827

24	104	100.0	527	8	ADO63826
25	104	100.0	527	9	ADZ19004
26	104	100.0	530	2	AAV34173
27	104	100.0	532	2	AAV17083
28	104	100.0	543	2	AAV02345
29	104	100.0	543	2	AAV17082
30	104	100.0	543	3	AAV57590
31	104	100.0	543	3	AAV57590
32	104	100.0	543	3	AAV52990
33	104	100.0	543	4	AAV97635
34	104	100.0	543	4	AAV97635
35	104	100.0	543	4	AAV97635
36	104	100.0	543	5	ABE07813
37	104	100.0	543	7	ADD18950
38	104	100.0	543	7	ADG88800
39	104	100.0	543	8	ADL16379
40	104	100.0	543	8	ADK52086
41	104	100.0	543	8	ADM48716
42	104	100.0	543	8	ADM48759
43	104	100.0	543	8	ADN05074
44	104	100.0	543	8	ADN04902
45	104	100.0	543	8	ADO63831
46	104	100.0	543	8	ADO63824
47	104	100.0	543	8	ADO63823
48	104	100.0	543	8	ADO63832
49	104	100.0	543	8	ADO63822
50	104	100.0	543	8	ADQ80372
51	104	100.0	543	8	ADR88210
52	104	100.0	543	8	ADP25079
53	104	100.0	543	8	ADT78177
54	104	100.0	543	9	ADY27036
55	104	100.0	543	9	ADY63087
56	104	100.0	543	9	AEA42466
57	104	100.0	543	9	AEA42426
58	104	100.0	543	10	ABE96848
59	104	100.0	545	6	ABP56822
60	104	100.0	545	7	ADE16012
61	104	100.0	545	8	ADL93951
62	104	100.0	556	9	ADZ19010
63	104	100.0	570	9	ADZ19008
64	104	100.0	588	2	AAV30124
65	104	100.0	592	2	AAV02346
66	104	100.0	592	3	ABO88850
67	104	100.0	592	7	ADG88804
68	104	100.0	592	8	ADL16383
69	104	100.0	592	8	ADM48720
70	104	100.0	592	9	AEA42461
71	99	95.2	22	9	ADY27042
72	99	95.2	170	8	ADL16424
73	99	95.2	170	8	ADM48758
74	99	95.2	174	9	ADZ19013
75	99	95.2	535	3	ABO88851
76	99	95.2	535	5	ABO87811
77	99	95.2	535	7	ADG88834
78	99	95.2	535	8	ADL16413
79	99	95.2	535	8	ADM48750
80	99	95.2	535	8	ADR88208
81	99	95.2	535	8	ADT78175
82	99	95.2	535	9	ADY27033
83	99	95.2	535	9	AEA42424
84	99	95.2	536	5	ABO87812
85	99	95.2	536	8	ADR88209
86	99	95.2	536	8	ADT78176
87	99	95.2	536	9	ADY27035
88	99	95.2	536	9	AEA42425
89	92	88.5	174	9	ADZ19014
90	92	88.5	545	9	ADY27034
91	81	77.9	15	9	ADU70836
92	81	77.9	15	9	ADU71279
93	81	77.9	173	9	ADZ19015
94	81	77.9	523	5	ABO87814
95	81	77.9	523	7	ABW02017
96	81	77.9	523	8	ADR88211

ADO63826	Chimeric
Adz19004	HepGS4 co
Ray34173	Human pre
AY17083	Seq ID No
AY02345	A human h
AY17082	Human hep
AY57590	Human hep
AY57590	Human hep
AB08849	Amino aci
AY52990	Human hep
AY97635	Human hep
AB86206	Human hep
AB88361	Human mem
AB07813	Human hep
ADD18950	Human dis
ADG88800	Human hpa
ADL16379	Human hep
ADK52086	Human ato
ADM48716	Human hpa
ADM48759	Human hpa
ADN05074	Antipsori
ADN04902	Antipsori
ADO63831	Human hep
ADO63824	Human hep
ADO63823	Human hep
ADO63832	Human hep
ADO63822	Human hep
ADQ80372	Heparanas
ADR88210	Human pre
ADP25079	PRO poly
ADT78177	Human hep
ADY27036	Human hep
ADY63087	Human clo
AEA42466	Human hep
AEA42426	Human hep
ABE96848	Human hep
ABP56822	Human hep
ADE16012	G-coupled
ADL93951	Human G-c
ADZ19010	Heparanas
ADZ19008	Heparanas
AAV30124	A human p
AAV02346	A human h
ABO88850	Amino aci
ADG88804	Human SK-
ADL16383	Human hep
ADM48720	Human SK-
AEA42461	Human hep
ADY27042	Heparanas
ADL16424	Partial r
ADM48758	Rat hpa p
ADZ19013	Murine he
ABO88851	A murine
ABO87811	Mouse hep
ADG88834	Mouse hpa
ADL16413	Mouse hep
ADM48750	Mouse hpa
ADR88208	Mouse hep
ADT78175	Mouse hep
ADY27033	Murine he
AEA42424	Mouse hep
ABO87812	Rat hepar
ADR88209	Rat hepar
ADT78176	Rat hepar
ADY27035	Rat hepar
AEA42425	Rat hepar
ADZ19014	Bovine he
ADY27034	Bovine he
ADU70836	Human hep
ADU71279	Human hep
ADZ19015	Chicken h
ABO87814	Chicken h
ABW02017	Chicken h
ADR88211	Chicken h



97	81	77.9	523	8	ADT78178	Adt78178	Chicken h	170	43	41.3	350	8	ADJ34798	Adj34798	Xylanase
98	81	77.9	523	9	ADY27037	Ady27037	Chicken h	171	43	41.3	1128	3	AAG29155	Aag29155	Arabidops
99	81	77.9	523	9	AEA42427	Aea42427	Chicken h	172	43	41.3	1206	3	AAG29154	Aag29154	Arabidops
100	80	76.9	15	9	ADU70951	Adu70951	Human hep	173	43	41.3	1293	3	AAG29153	Aag29153	Arabidops
101	77	74.0	15	9	ADU71167	Adu71167	Human hep	174	42.5	40.9	227	8	ADX66278	Adx66278	Plant ful
102	76	73.1	15	9	ADU70835	Adu70835	Human hep	175	42.5	40.9	314	8	ADX66660	Adx66660	Plant ful
103	72	69.2	15	9	ADU71278	Adu71278	Human hep	176	42.5	40.9	690	8	ADX72696	Adx72696	Plant ful
104	71	68.2	15	9	ADU71168	Adu71168	Human hep	177	42.5	40.9	770	2	AAy43478	Aay43478	Amino aci
105	66	63.5	15	9	ADU71012	Adu71012	Human hep	178	42	40.4	206	6	ABP80884	Abp80884	N. gonorr
106	66	63.5	32	2	AAy34175	Aay34175	Human pre	179	42	40.4	241	9	AEA22109	Aea22109	Campyloba
107	66	63.5	189	9	ADZ19016	Adz19016	Heparanas	180	42	40.4	305	3	AAG33465	Aag33465	Arabidops
108	66	63.5	282	4	AAm24147	Aam24147	Human EST	181	42	40.4	326	3	AAG33464	Aag33464	Arabidops
109	66	63.5	439	4	AAU07423	Aau07423	Human hep	182	42	40.4	404	8	ADV88147	Adv88147	Streptoco
110	66	63.5	480	4	AAy97634	Aay97634	Human hep	183	42	40.4	404	8	ADV79400	Adv79400	Streptoco
111	66	63.5	480	4	AAU07418	Aau07418	Novel hum	184	42	40.4	418	3	AAG33463	Aag33463	Arabidops
112	66	63.5	480	4	AAU07418	Aau07418	Novel hum	185	42	40.4	428	5	ABG93810	Abg93810	Herbicida
113	66	63.5	482	4	AAU07418	Aau07418	Novel hum	186	42	40.4	429	8	ADT55647	Adt55647	Plant pol
114	66	63.5	534	4	AAU07424	Aau07424	Human hep	187	42	40.4	448	8	ABO58787	Abos8787	Human gen
115	66	63.5	534	4	AAU07424	Aau07424	Human hep	188	42	40.4	505	5	ABP30486	Abp30486	Streptoco
116	66	63.5	534	5	AAU07424	Aau07424	Human hep	189	42	40.4	505	5	ABP30486	Abp30486	Streptoco
117	66	63.5	534	5	AAU07424	Aau07424	Human hep	190	42	40.4	518	5	ABP27721	Abp27721	Streptoco
118	66	63.5	534	4	AAU07424	Aau07424	Human hep	191	42	40.4	518	5	ABP27721	Abp27721	Streptoco
119	66	63.5	534	4	AAU07424	Aau07424	Human hep	192	42	40.4	742	8	ADS10736	Ads10736	Human ova
120	66	63.5	534	4	AAU07424	Aau07424	Human hep	193	42	40.4	772	8	ADP23017	Adp23017	PRO polyp
121	66	63.5	534	4	AAU07424	Aau07424	Human hep	194	42	40.4	772	8	ADP23017	Adp23017	PRO polyp
122	66	63.5	534	4	AAU07424	Aau07424	Human hep	195	42	40.4	791	9	AED74134	Aed74134	Human pla
123	61	58.7	470	5	AAU07424	Aau07424	Human hep	196	42	40.4	856	2	AAR96949	Aar96949	Phosphoen
124	61	58.7	528	5	AAU07424	Aau07424	Human hep	197	41.5	39.9	111	9	AED02424	Aed02424	T84.66 an
125	61	58.7	528	5	AAU07424	Aau07424	Human hep	198	41.5	39.9	127	8	ABO60044	Abos60044	Human gen
126	60	57.7	15	9	ADU71166	Adu71166	Human hep	199	41.5	39.9	132	2	AAR10920	Aar10920	kappa lig
127	59	56.7	15	9	ADU71014	Adu71014	Human hep	200	41.5	39.9	145	5	ABG98792	Abg98792	Human rib
128	59	56.7	98	5	ADU71014	Adu71014	Human hep	201	41.5	39.9	179	8	ADS17745	Ads17745	Anti-CEA
129	55	52.9	15	9	ADU70950	Adu70950	Human hep	202	41.5	39.9	184	4	AAM38805	Aam38805	Human pol
130	54	51.9	15	9	ADU70913	Adu70913	Human hep	203	41.5	39.9	184	4	AAB60105	Aab60105	Human tra
131	50	48.1	9	9	ADU70628	Adu70628	Human hep	204	41.5	39.9	184	8	ADN05540	Adn05540	Antipsori
132	50	48.1	9	9	ADU70782	Adu70782	Human hep	205	41.5	39.9	184	8	ADN05540	Adn05540	Antipsori
133	50	48.1	147	9	ADZ19018	Adz19018	Heparanas	206	41.5	39.9	184	8	ADP55889	Adp55889	Human PRO
134	50	48.1	488	4	AAU07424	Aau07424	Human hep	207	41.5	39.9	184	9	ADY19167	Ady19167	PRO polyp
135	50	48.1	488	4	AAU07424	Aau07424	Human hep	208	41.5	39.9	196	4	AAU40591	Aam40591	Human pol
136	50	48.1	488	4	AAU07424	Aau07424	Human hep	209	41.5	39.9	262	9	AED84408	Aed84408	Murine mo
137	49	47.1	9	2	AAU07424	Aau07424	Human hep	210	41.5	39.9	623	9	AED84408	Aed84408	CAB molec
138	49	47.1	9	9	ADU70440	Adu70440	Human hep	211	41.5	39.9	710	8	ADS42585	Ads42585	Bacterial
139	49	47.1	9	9	ADU70571	Adu70571	Human hep	212	41.5	39.9	1040	8	ADS17749	Ads17749	Rat major
140	49	47.1	9	9	ADU70351	Adu70351	Human hep	213	41.5	39.9	1072	8	ADS17747	Ads17747	Human maj
141	49	47.1	9	9	ADU70811	Adu70811	Human hep	214	41.5	39.9	1127	8	ADS17789	Ads17789	Polylysin
142	49	47.1	15	9	ADU70856	Adu70856	Human hep	215	41.5	39.9	1159	8	ADS17787	Ads17787	Polylysin
143	48	46.2	9	9	ADU70535	Adu70535	Human hep	216	41.5	39.9	1206	8	ADS17763	Ads17763	GAL4 + ra
144	48	46.2	9	9	ADU70629	Adu70629	Human hep	217	41.5	39.9	1238	8	ADS17761	Ads17761	GAL4 + hu
145	48	46.2	151	9	ADZ19017	Adz19017	Heparanas	218	41.5	39.9	1273	8	ADS17777	Ads17777	MS2 pepti
146	48	46.2	261	4	AAU07424	Aau07424	Human hep	219	41	39.4	116	8	ADL05804	Adl05804	M. catarr
147	48	46.2	261	4	AAU07424	Aau07424	Human hep	220	41	39.4	138	9	ADZ19021	Adz19021	Heparanas
148	48	46.2	279	2	AAU07424	Aau07424	Human hep	221	41	39.4	274	8	ADX8300	Adx8300	Plant ful
149	48	46.2	411	5	ABP66200	Abp66200	Bifidobac	222	41	39.4	274	8	ADY12676	Ady12676	Plant ful
150	47	45.2	9	9	ADU70684	Adu70684	Human hep	223	41	39.4	278	8	ADX80584	Adx80584	Plant ful
151	47	45.2	9	9	ADU70594	Adu70594	Human hep	224	41	39.4	281	8	ADX66918	Adx66918	Plant ful
152	47	45.2	9	9	ADU70685	Adu70685	Human hep	225	41	39.4	333	5	ABP27264	Abp27264	Streptoco
153	46	44.2	150	9	ADZ19019	Adz19019	Heparanas	226	41	39.4	342	3	AAG14301	Aag14301	Arabidops
154	46	44.2	414	4	ABG25051	Abg25051	Novel hum	227	41	39.4	346	3	AAG14301	Aag14301	Arabidops
155	46	44.2	488	4	AAU07424	Aau07424	Human hep	228	41	39.4	351	6	ABM70737	Abm70737	Photorhab
156	45	43.3	9	9	ADU70683	Adu70683	Human hep	229	41	39.4	367	2	ADT58411	Adt58411	Plant pol
157	45	43.3	486	8	ADU07424	Adu07424	Human hep	230	41	39.4	412	3	AAG14299	Aag14299	Arabidops
158	45	43.3	3540	8	ADN96836	Adn96836	Bugula br	231	41	39.4	442	8	ADX66334	Adx66334	Plant ful
159	44	42.3	9	9	ADU70686	Adu70686	Human hep	232	41	39.4	694	9	ABM91425	Abm91425	M. xanthu
160	44	42.3	15	9	ADU70952	Adu70952	Human hep	233	41	39.4	754	3	ABJ35987	Abj35987	Sorbitol
161	44	42.3	87	4	AAU55506	Aau55506	Propionib	234	41	39.4	848	2	AAJ34512	Aaj34512	Porphorym
162	44	42.3	87	6	ABM52025	Abm52025	Propionib	235	41	39.4	878	2	AAJ34385	Aaj34385	Porphorym
163	44	42.3	383	8	ADN96836	Adn96836	Bugula br	236	41	39.4	5069	2	AAG52846	Aag52846	A. medite
164	44	42.3	569	8	ADN96836	Adn96836	Bugula br	237	40	38.5	76	7	ABO72337	Abos72337	Pseudomon
165	44	42.3	659	1	AAU07424	Aau07424	Human hep	238	40	38.5	96	3	ABJ41008	Abj41008	Human ORF
166	44	42.3	779	5	ABM07845	Abm07845	Human Nap	239	40	38.5	96	5	ABP34610	Abp34610	Human ORF
167	44	42.3	997	8	ADX66509	Adx66509	Plant ful	240	40	38.5	113	4	AAO06297	Aao06297	Human pol
168	44	42.3	998	9	ADU05670	Adu05670	Cyclin-de	241	40	38.5	158	6	ADA34928	Ada34928	Acinetoba
169	43	41.3	253	6	ABU48538	Abu48538	Protein e	242	40	38.5	178	5	ABB48544	Abb48544	Listeria

243	40	38.5	178	6	ABU32420	Abu32420 Protein e	316	39	37.5	113	4	AAU31556	Aau31556 Novel hum
244	40	38.5	204	4	AAU19298	Aau19298 Human G p	317	39	37.5	119	4	AAO03948	Aao03948 Human pol
245	40	38.5	211	4	AGS90850	Agsg90850 C glutami	318	39	37.5	127	2	AAU12885	Aay12885 Human 5'
246	40	38.5	240	3	AAU57080	Abbs57080 Human pro	319	39	37.5	149	4	ABG23456	ABG23456 Novel hum
247	40	38.5	262	4	ABB66342	Abbs66342 Drosophil	320	39	37.5	150	5	ABP07825	Abp07825 Human ORF
248	40	38.5	287	4	ABB66343	Abbs66343 Drosophil	321	39	37.5	158	8	ABM70650	Abm70650 Photorhab
249	40	38.5	297	3	AAU15814	Aau15814 Arabidops	322	39	37.5	225	6	ABW70650	Abw70650 Thermoco
250	40	38.5	299	7	ADP40114	Adp40114 Human NOV	323	39	37.5	227	3	AAO70727	Aao70727 Arabidops
251	40	38.5	302	4	ABB69502	Abbs69502 Drosophil	324	39	37.5	235	4	AAO63451	Aao63451 Human bre
252	40	38.5	326	4	ABB71638	Abbs71638 Drosophil	325	39	37.5	235	4	AAO63280	Aao63280 Human bre
253	40	38.5	331	3	AAU95072	Aau95072 Candida a	326	39	37.5	241	8	ADU71344	Adu71344 Plant ful
254	40	38.5	333	8	ADU88472	Adu88472 Streptoco	327	39	37.5	279	7	ADE12789	Adel2789 L. rhanno
255	40	38.5	333	8	ADU79725	Adv79725 Streptoco	328	39	37.5	281	3	AAO70726	Aao70726 Arabidops
256	40	38.5	333	8	ADU81883	Adv81883 Streptoco	329	39	37.5	281	7	ADP88603	Adp88603 Ribosome
257	40	38.5	356	6	ABG73525	Abg73525 B. subtil	330	39	37.5	281	9	ADU34638	Adv34638 Des-val a
258	40	38.5	362	8	ADU71757	Ado71757 Human sel	331	39	37.5	298	4	ABG06306	Abg06306 Novel hum
259	40	38.5	362	8	ADP90527	Adp90527 Human sel	332	39	37.5	298	7	ADF60292	Adf60292 Human con
260	40	38.5	364	8	ADU77474	Adu77474 Plant ful	333	39	37.5	298	7	ADF58916	Adf58916 Human pol
261	40	38.5	365	2	AAU47874	Aat47874 Enzyme/bi	334	39	37.5	298	7	ADF60477	Adf60477 Human con
262	40	38.5	365	2	AAU09195	Aau09195 Rhodococc	335	39	37.5	305	8	ADN47029	Adn47029 Thermoco
263	40	38.5	365	2	AAW97239	Aaw97239 An enzyme	336	39	37.5	313	8	ADU07646	Ady07646 Plant ful
264	40	38.5	365	2	AAW89242	Aaw89242 Rhodococc	337	39	37.5	313	8	ADU93718	Adx93718 Plant ful
265	40	38.5	365	4	AAU02430	Aau02430 Rhodococc	338	39	37.5	319	3	AAO70725	Aao70725 Arabidops
266	40	38.5	365	7	AAE37615	Aae37615 Rhodococc	339	39	37.5	326	7	ABO75715	Abg075715 Pseudomon
267	40	38.5	373	3	AAU15813	Aau15813 Arabidops	340	39	37.5	338	4	ABG23449	Abg23449 Novel hum
268	40	38.5	380	7	ADN95151	Adn95151 Human BEC	341	39	37.5	340	4	AAU04820	Aau04820 Micromono
269	40	38.5	381	3	AAO03188	Aao03188 Human sel	342	39	37.5	340	6	ABP99240	Abp99240 Orthosomy
270	40	38.5	381	5	AAU84306	Aau84306 Human end	343	39	37.5	342	7	ABO76674	Abg076674 Pseudomon
271	40	38.5	381	6	ABU89753	Abu89753 Protein d	344	39	37.5	347	8	ADU70642	Adx70642 Plant ful
272	40	38.5	381	7	ADD47226	Ada47226 Human Pro	345	39	37.5	355	8	ADU57590	Adt57590 Plant pol
273	40	38.5	381	7	ADP65180	Adp65180 Human sel	346	39	37.5	362	6	AAO16382	Aao16382 Drosophil
274	40	38.5	381	7	ABE86458	Aeb86458 Human sel	347	39	37.5	373	8	ADS27581	Ads27581 Bacterial
275	40	38.5	381	8	ADU69218	Adu69218 Human sel	348	39	37.5	376	8	ADU09636	Ady09636 Plant ful
276	40	38.5	381	9	AEC12741	Aec12741 Human sur	349	39	37.5	385	9	ADD47224	Adx47224 Rat Prote
277	40	38.5	381	9	AEC12381	Aec12381 Human sur	350	39	37.5	385	9	AEC12594	Aec12594 Rat surro
278	40	38.5	381	10	AE88077	Aee88077 Human pro	351	39	37.5	385	9	AEC12061	Aec12061 Rat surro
279	40	38.5	404	3	AG15812	Agl5812 Arabidops	352	39	37.5	387	4	ABU52973	Abu52973 Human tes
280	40	38.5	404	8	ADU55977	Adt55977 Plant pol	353	39	37.5	391	5	ABP62776	Abp62776 Protein f
281	40	38.5	477	8	ADU43285	Adt43285 Bacterial	354	39	37.5	391	7	ADJ72187	Adj72187 S roseosp
282	40	38.5	512	9	ABM92453	Abm92453 M. xanthu	355	39	37.5	394	4	AAO89139	Adg89139 Human sec
283	40	38.5	546	5	ABG93260	Abg93260 C. albica	356	39	37.5	394	7	ADM05625	Adm05625 Human pro
284	40	38.5	546	5	ABG93297	Abg93297 C. albica	357	39	37.5	394	7	AEC88555	Aec88555 Human CDN
285	40	38.5	550	5	ABG93259	Abg93259 C. albica	358	39	37.5	404	7	ADG48207	Adg48207 Saccharom
286	40	38.5	550	5	ABG93311	Abg93311 C. albica	359	39	37.5	404	9	AEC55897	Aec55897 Saccharom
287	40	38.5	550	5	ABG93365	Abg93365 Human BAX	360	39	37.5	417	4	ABB71426	Abb71426 Drosophil
288	40	38.5	580	4	AAU90361	Aau90361 C glutami	361	39	37.5	419	4	ABG66118	Abg66118 Drosophil
289	40	38.5	585	8	ADN22546	Adn22546 Bacterial	362	39	37.5	423	4	ABG60949	Abg60949 Drosophil
290	40	38.5	633	4	AAU79283	Abu79283 Corynebac	363	39	37.5	426	8	ABE86882	Aeb86882 Polyketid
291	40	38.5	637	4	ABG08572	Abg08572 Novel hum	364	39	37.5	426	8	ABE86679	Aeb86679 Polyketid
292	40	38.5	637	4	ABG03188	Abg03188 Novel hum	365	39	37.5	441	7	ADG28432	Adg28432 Soybean S
293	40	38.5	701	8	ADS30240	Ads30240 Bacterial	366	39	37.5	450	7	ADG08445	Adg08445 Novel pro
294	40	38.5	713	9	ABM97151	Abm97151 M. xanthu	367	39	37.5	539	8	ADY13296	Ady13296 Plant ful
295	40	38.5	742	2	AAU05235	Aau05235 Amino aci	368	39	37.5	561	5	ABG97480	Abg97480 Human NOV
296	40	38.5	743	4	ABG23278	Abg23278 Novel hum	369	39	37.5	651	3	AAU10936	Aau10936 Human RNA
297	40	38.5	743	4	ABG14502	Abg14502 Novel hum	370	39	37.5	651	3	ABU78308	Abu78308 Amino aci
298	40	38.5	7968	10	ABE68641	Aee68641 Streptomy	371	39	37.5	661	8	ADN24139	Adn24139 Bacterial
299	39.5	38.0	388	2	AAU32058	Aay32058 Cat pregn	372	39	37.5	730	9	AEA49153	Aea49153 L. rhanno
300	39.5	38.0	388	7	ADF11571	Adf11571 Feline pr	373	39	37.5	730	9	AED03561	Aed03561 L. rhanno
301	39.5	38.0	551	8	ADS23997	Ads23997 Bacterial	374	39	37.5	735	7	ABM86341	Abm86341 Rice abio
302	39.5	38.0	584	8	ADU04704	Adi04704 M. catarr	375	39	37.5	756	8	ADU91671	Adx91671 Plant ful
303	39	37.5	9	9	ADU70627	Adu70627 Human hep	376	39	37.5	782	7	ADF93177	Adf93177 Human N-a
304	39	37.5	15	9	ADU171011	Adu171011 Human hep	377	39	37.5	782	7	ADF93183	Adf93183 Human N-a
305	39	37.5	31	4	AAU18407	Aam18407 Peptide #	378	39	37.5	784	7	ADF93185	Adf93185 Human N-a
306	39	37.5	31	4	ABB37446	Abb37446 Peptide #	379	39	37.5	791	7	ADF93179	Adf93179 Murine N-
307	39	37.5	31	4	AAU30883	Aam30883 Peptide #	380	39	37.5	792	8	ADO28695	Ado28695 Human can
308	39	37.5	31	4	ABB32197	Abb32197 Peptide #	381	39	37.5	792	8	ADS32728	Ads32728 Human N-a
309	39	37.5	31	4	ABB22742	Abb22742 Protein #	382	39	37.5	792	8	ADU05159	Adu05159 Human N-a
310	39	37.5	31	4	AAU70567	Aam70567 Human bon	383	39	37.5	821	5	AAE25023	Aee25023 Human dru
311	39	37.5	31	4	AAU58124	Aam58124 Human bra	384	39	37.5	875	8	ADS28775	Ads28775 Bacterial
312	39	37.5	31	4	ABG52248	Abg52248 Human liv	385	39	37.5	909	3	AAU94328	Aay94328 Maize DNA
313	39	37.5	31	4	AAU06007	Aam06007 Peptide #	386	39	37.5	1360	8	ADO26836	Ado26836 Human rec
314	39	37.5	31	5	ABG40231	Abg40231 Human pep	387	39	37.5	5432	8	ADQ91708	Adq91708 Polyketid
315	39	37.5	67	4	ABG23453	Abg23453 Novel hum	388	39	37.5	5432	8	ABE87000	Aeb87000 Streptomy

389	39	5432	8	ABE86797	AbB86797 Streptomy	462	38	36.5	503	7	ABO69145	AbO69145 Pseudom
390	38.5	67	3	ADA32924	Ada32924 Acinetoba	463	38	36.5	525	6	ADA56830	Ada56830 Human sec
391	38.5	120	3	AY56701	Ray56701 Rat anti-	464	38	36.5	525	6	ABr47695	ABr47695 Human sec
392	38.5	186	7	ABM87982	ABm87982 Rice abio	465	38	36.5	525	6	ABR00046	ABr00046 Human gen
393	38.5	322	6	ADBI7492	ADbi7492 Soybean p	466	38	36.5	525	7	ADB91501	ADB91501 Human sec
394	38.5	322	9	REC75716	Rec75716 Soybean A	467	38	36.5	525	7	ADC74074	ADC74074 Human sec
395	38.5	397	8	ADN46729	ADn46729 Thermoco	468	38	36.5	530	4	AB57868	AB57868 Drosophi
396	38.5	408	3	AY87978	AY87978 Pseudomoc	469	38	36.5	540	8	ADQ89648	ADQ89648 Antagonis
397	38.5	539	8	ADN17319	ADn17319 D. destru	470	38	36.5	553	8	ADN20097	ADn20097 Bacterial
398	38.5	574	3	AY87977	AY87977 Pseudomon	471	38	36.5	573	2	AAW48874	AAw48874 Hyphozyma
399	38.5	667	3	AY87976	AY87976 Pseudomon	472	38	36.5	577	6	ABU32269	ABu32269 Protein e
400	38.5	984	5	AB93326	AB93326 Herbicida	473	38	36.5	582	7	ABO66203	ABO66203 Klebsiell
401	38.5	1545	9	ADZ44910	ADz44910 D carota	474	38	36.5	599	6	ADA33364	ADA33364 Acinetoba
402	38	36.5	9	ADU70466	ADu70466 Human hep	475	38	36.5	602	8	AAW47598	AAw47598 Drosophi
403	38	36.5	15	ADU71269	ADu71269 Human hep	476	38	36.5	692	8	ADS42094	ADS42094 Bacterial
404	38	36.5	20	AY338490	AY338490 Human sec	477	38	36.5	738	2	AAr13393	AAr13393 A. altocet
405	38	36.5	45	ABB41509	ABb41509 Peptide #	478	38	36.5	738	2	AAr20192	AAr20192 ADH compl
406	38	36.5	45	AAW35301	AAw35301 Peptide #	479	38	36.5	761	4	AAW46719	AAw46719 I. brucei
407	38	36.5	45	ABW25382	ABW25382 Protein #	480	38	36.5	785	8	ADL98278	ADL98278 Human ins
408	38	36.5	45	AAW75189	AAw75189 Human bon	481	38	36.5	785	9	ADW00634	ADW00634 Human pro
409	38	36.5	45	AAW62379	AAw62379 Human bra	482	38	36.5	834	4	AAW61238	AAw61238 Murine M-
410	38	36.5	45	ABG56951	ABg56951 Human liv	483	38	36.5	834	6	ABO32675	ABO32675 Secreted
411	38	36.5	45	ABG44852	ABg44852 Human pep	484	38	36.5	834	7	AB90780	AB90780 Mouse M-S
412	38	36.5	78	ABW17364	ABW17364 Human ner	485	38	36.5	834	7	ADP71515	ADP71515 Murine M-
413	38	36.5	131	AAW30278	AAw30278 Light cha	486	38	36.5	834	8	ADQ10333	ADQ10333 Human pol
414	38	36.5	138	ADZ19020	ADz19020 Heparanas	487	38	36.5	834	8	ADR67091	ADR67091 Mouse can
415	38	36.5	170	AG74676	AG74676 Human col	488	38	36.5	856	6	ADA55361	ADA55361 Human pro
416	38	36.5	217	ADS27809	ADs27809 Bacterial	489	38	36.5	883	8	ADI45397	ADI45397 Rice isop
417	38	36.5	225	AAW33478	AAw33478 Arabidops	490	38	36.5	896	8	ADW76298	ADW76298 Plant ful
418	38	36.5	238	ABM96838	ABm96838 M. xanthu	491	38	36.5	921	9	AEF53848	AEf53848 Human ins
419	38	36.5	240	ABM96678	ABm96678 M. xanthu	492	38	36.5	921	9	AEF53848	AEf53848 Human ins
420	38	36.5	247	AG13477	AG13477 Arabidops	493	38	36.5	924	5	ABW54116	ABW54116 Lactococ
421	38	36.5	233	AG43712	AG43712 Arabidops	494	38	36.5	924	8	ADS29350	ADS29350 Bacterial
422	38	36.5	255	ADS21205	ADs21205 Bacterial	495	38	36.5	1058	6	ABU26274	ABu26274 Protein e
423	38	36.5	257	AG43711	AG43711 Arabidops	496	38	36.5	1079	5	ABP73500	ABP73500 Candida a
424	38	36.5	273	ABR62846	ABr62846 Human car	497	38	36.5	1141	7	ABM89610	ABm89610 Rice abio
425	38	36.5	274	AAW38486	AAw38486 Human sec	498	38	36.5	1157	8	ADP99144	ADP99144 Human tra
426	38	36.5	276	ADW78137	ADw78137 Plant ful	499	38	36.5	1213	8	ADP98964	ADP98964 C. albica
427	38	36.5	278	ABU43175	ABu43175 Protein e	500	38	36.5	1268	4	ABW11702	ABW11702 Human ins
428	38	36.5	279	AG43710	AG43710 Arabidops	501	38	36.5	1268	4	ABW11702	ABW11702 Human ins
429	38	36.5	279	ABP39377	ABp39377 Staphyloc	502	38	36.5	1268	7	ADQ09341	ADQ09341 Novel pro
430	38	36.5	279	ADS04554	ADs04554 Staphyloc	503	38	36.5	1282	5	AAE19143	AAE19143 Human kin
431	38	36.5	284	ABU21840	ABu21840 Protein e	504	38	36.5	1297	3	AAW36840	AAW36840 Human ins
432	38	36.5	341	ADN22890	ADn22890 Bacterial	505	38	36.5	1297	9	AEF53847	AEf53847 Human ins
433	38	36.5	341	ADN22896	ADn22896 Bacterial	506	38	36.5	1310	6	ABP96077	ABP96077 Human pro
434	38	36.5	343	ADW1554	ADW1554 Plant ful	507	38	36.5	1374	8	ADN22758	ADn22758 Bacterial
435	38	36.5	353	ADW48962	ADW48962 Rat Prote	508	38	36.5	1400	7	ADL33323	ADL33323 Human tra
436	38	36.5	353	ADW48170	ADW48170 Rat Prote	509	38	36.5	1418	4	ABG02184	ABG02184 Novel hum
437	38	36.5	359	AAW97868	AAw97868 Hamster p	510	38	36.5	1435	10	AAE72192	AAE72192 Human tar
438	38	36.5	359	ADE60176	ADe60176 Human pro	511	38	36.5	1440	7	ADE08390	ADE08390 Novel REM
439	38	36.5	359	ADE60174	ADe60174 Rat Prote	512	38	36.5	1469	6	ABR43185	ABr43185 Human REM
440	38	36.5	359	ADA48964	ADa48964 Human pro	513	38	36.5	1469	8	ABM84440	ABm84440 Human dia
441	38	36.5	359	ADA48172	ADa48172 Human pro	514	38	36.5	1503	3	AAW92944	AAW92944 Human TRP
442	38	36.5	359	ABO66595	ABo66595 Klebsiell	515	38	36.5	1503	3	AAW95439	AAW95439 Human cal
443	38	36.5	359	ADW57715	ADw57715 Rheumatoi	516	38	36.5	1503	3	AAW36865	AAW36865 Human put
444	38	36.5	359	ADY15754	ADy15754 PRO polyp	517	38	36.5	1503	5	ABW76459	ABW76459 Human lon
445	38	36.5	359	ABP79596	ABP79596 N. gonorr	518	38	36.5	1503	5	ABW84544	ABW84544 Human tra
446	38	36.5	361	ABW91031	ABW91031 Herbicida	519	38	36.5	1503	7	ADC47022	ADC47022 Human LTR
447	38	36.5	364	ADT58959	ADt58959 Plant pol	520	38	36.5	1503	7	ADC77685	ADC77685 Human 222
448	38	36.5	406	ADW9028	ADw9028 Human pro	521	38	36.5	1503	7	ADC83633	ADC83633 LTRPC3-re
449	38	36.5	406	AEAL9966	AEa19966 Novel hum	522	38	36.5	1503	8	ADQ89102	ADQ89102 Human uro
450	38	36.5	409	ABU00700	ABu00700 S. pneumo	523	38	36.5	1503	8	ADT93474	ADt93474 Human tra
451	38	36.5	409	ABP81316	ABp81316 Streptoco	524	38	36.5	1503	9	ABW85144	ABW85144 Human nov
452	38	36.5	409	ABU45740	ABu45740 Protein e	525	38	36.5	1583	5	ABP70147	ABP70147 Human NOV
453	38	36.5	414	ABU36068	ABu36068 Protein e	526	38	36.5	1585	5	AAE21186	AAE21186 Human TRI
454	38	36.5	414	ADW74341	ADw74341 Mycobacte	527	38	36.5	1619	6	ADA36060	ADA36060 Acinetoba
455	38	36.5	426	ABE86878	ABe86878 Polyketid	528	38	36.5	1624	5	ABB81576	ABB81576 Human ATP
456	38	36.5	426	ABE86675	ABe86675 Polyketid	529	38	36.5	1624	5	ABB81582	ABB81582 Human ABC
457	38	36.5	435	ABU40192	ABu40192 Protein e	530	38	36.5	1624	6	ABB84638	ABB84638 Human ABC
458	38	36.5	440	AAU33635	AAU33635 Pseudomon	531	38	36.5	1624	10	AAE72191	AAE72191 Human tar
459	38	36.5	440	ABU38545	ABu38545 Protein e	532	38	36.5	1808	10	AEF57853	AEf57853 Polyketid
460	38	36.5	461	ABU20543	ABu20543 Protein e	533	38	36.5	2088	10	AEF16945	AEf16945 Streptomy
461	38	36.5	469	ABM88848	ABm88848 Rice abio	534	38	36.5	5207	8	ADQ91706	ADQ91706 Polyketid

535	38	36.5	5207	8	AE8B6998	Aeb86998 Streptomy	608	37	35.6	389	2	AAW53158	Aaw53158 Arthrobac
536	38	36.5	5207	8	AE8B6795	Aeb86795 Streptomy	609	37	35.6	389	2	AAW53162	Aaw53162 Arthrobac
537	38	36.5	5712	10	AE8B8639	Aee86839 Streptomy	610	37	35.6	389	2	AAW53161	Aaw53161 Arthrobac
538	38	36.5	7257	3	AA558576	Aay58576 Sorangium	611	37	35.6	389	2	AAW53164	Aaw53164 Arthrobac
539	37.5	36.1	72	3	AAQ02738	Aeq02738 Human sec	612	37	35.6	389	2	AAW53160	Aaw53160 Arthrobac
540	37.5	36.1	123	7	ADL96646	Adl96646 M. paratu	613	37	35.6	389	4	AAW53160	Aaw53160 Arthrobac
541	37.5	36.1	227	5	ABG70073	Abg70073 Human pre	614	37	35.6	389	8	ADO23559	Ado23559 Arthrobac
542	37.5	36.1	246	7	ADC87249	Adc87249 Human GPC	615	37	35.6	390	1	AAW53160	Aaw53160 Arthrobac
543	37.5	36.1	262	9	AD84412	Ad84412 Modified	616	37	35.6	390	1	AAW53160	Aaw53160 Arthrobac
544	37.5	36.1	274	9	AD84412	Ad84412 Modified	617	37	35.6	390	1	AAW53160	Aaw53160 Arthrobac
545	37.5	36.1	347	4	AB859052	Ab859052 Drosophil	618	37	35.6	397	5	ABP66124	Abp66124 Bifidobac
546	37.5	36.1	375	5	AB857201	Ab857201 Mouse tcc	619	37	35.6	401	6	ABU32343	Abu32343 Protein e
547	37.5	36.1	375	9	AD84414	Ad84414 CAB molec	620	37	35.6	402	7	ABO67435	Ab067435 Klebsiell
548	37.5	36.1	411	7	ADC30913	Adc30913 Human nov	621	37	35.6	403	8	ADT58194	Adt58194 Plant pol
549	37.5	36.1	430	9	ACA89042	Aca89042 Chicken t	622	37	35.6	405	8	ADC96787	Adc96787 E. faeciu
550	37.5	36.1	623	9	AEA99706	Aea99706 Human CAB	623	37	35.6	410	4	ABB65321	Abb65321 Drosophil
551	37.5	36.1	623	9	AEA99706	Aea99706 Human CAB	624	37	35.6	411	7	ADC95952	Adc95952 Arthrobac
552	37.5	36.1	623	9	AEA99706	Aea99706 Human CAB	625	37	35.6	418	4	ABG29768	Abg29768 Novel hum
553	37.5	36.1	1296	2	AAW88448	Aaw88448 Caenorhab	626	37	35.6	420	8	ADP45524	Adp45524 Human col
554	37.5	36.1	1559	9	ADZ44912	Adz44912 L. esculen	627	37	35.6	425	4	ABE62474	Ab62474 Drosophil
555	37	35.6	31	2	AAW62083	Aaw62083 Hyphozyma	628	37	35.6	427	7	ADF05940	Adf05940 Bacterial
556	37	35.6	75	4	AAW86624	Aaw86624 Human imm	629	37	35.6	430	6	ADA34337	Ada34337 Acinetoba
557	37	35.6	135	8	ADM66293	Adm66293 Plant ful	630	37	35.6	432	6	ABU26569	Abu26569 Protein e
558	37	35.6	138	7	ADM04818	Adm04818 Human pro	631	37	35.6	438	6	ABM15924	Abm15924 Mycobacte
559	37	35.6	138	9	AE87748	Aec87748 Human CDN	632	37	35.6	439	6	ABU24183	Abu24183 Protein e
560	37	35.6	142	4	AAU55549	Aau55549 Propionib	633	37	35.6	442	8	ADX77733	Adx77733 Plant ful
561	37	35.6	142	6	ABM52068	Abm52068 Propionib	634	37	35.6	459	6	ABU20396	Abu20396 Protein e
562	37	35.6	149	6	AAU44605	Aau44605 Protein e	635	37	35.6	470	7	ABU80211	Abu80211 Mycobacte
563	37	35.6	151	4	AAU04698	Aau04698 Human pol	636	37	35.6	478	6	ABG74844	Abg74844 Potato 9-
564	37	35.6	211	1	AAPI0032	Api0032 Sequence	637	37	35.6	496	8	ADQ08798	Adq08798 Clona int
565	37	35.6	227	1	AAPI0035	Api0035 Sequence	638	37	35.6	505	2	AAU35699	Aay35699 C. pneumo
566	37	35.6	228	7	ADJ11584	Adj11584 Rice prot	639	37	35.6	516	5	ABY47838	Adh47838 Listeria
567	37	35.6	254	5	ABP28106	Abp28106 Streptoco	640	37	35.6	517	5	ADH87433	Adh87433 Enterococ
568	37	35.6	254	8	ADV88855	Adv88855 Streptoco	641	37	35.6	541	4	ABE62565	Ab62565 Drosophil
569	37	35.6	254	8	ADV82241	Adv82241 Streptoco	642	37	35.6	568	7	ADJ70002	Adj70002 Human hea
570	37	35.6	254	8	ADV80108	Adv80108 Streptoco	643	37	35.6	572	8	ADX73963	Adx73963 Plant ful
571	37	35.6	266	7	ADJ11794	Adj11794 Rice prot	644	37	35.6	578	8	ADN23662	Adn23662 Bacterial
572	37	35.6	266	7	ADJ11440	Adj11440 Rice prot	645	37	35.6	607	6	ABG72573	Abg72573 BSA abto
573	37	35.6	272	7	ABR42950	Ab42950 Human exo	646	37	35.6	607	6	ABO07192	Ab07192 Human p53
574	37	35.6	296	8	ADK46448	Adk46448 Streptoco	647	37	35.6	607	7	ADB63300	Ad63300 Human pro
575	37	35.6	306	7	ADF07209	Adf07209 Bacterial	648	37	35.6	607	8	ADQ95952	Adq95952 T cell ac
576	37	35.6	309	7	ABO63712	Ab063712 Klebsiell	649	37	35.6	607	8	ADQ95990	Adq95990 T cell ac
577	37	35.6	314	4	AAW79059	Aaw79059 Corynebac	650	37	35.6	638	8	ADU25696	Adu25696 L. acidop
578	37	35.6	314	4	AAW91752	Aag91752 C. glutami	651	37	35.6	638	9	AECS7426	Adc7426 L. acidop
579	37	35.6	314	7	ADD13753	Ad13753 C. glutam	652	37	35.6	684	8	ADG42109	Adg42109 Bacterial
580	37	35.6	319	7	ADM26740	Adm26740 Hyperther	653	37	35.6	695	7	ADB70236	Adb70236 C. neofo
581	37	35.6	331	7	ADM25756	Adm25756 Hyperther	654	37	35.6	700	6	ABU23181	Abu23181 Protein e
582	37	35.6	336	6	ABP81638	Abp81638 Streptoco	655	37	35.6	884	6	ABR43327	Ab43327 Dictyoglo
583	37	35.6	337	8	ABP67824	Abp67824 Drosophil	656	37	35.6	884	6	ABR43348	Ab43348 Dictyoglo
584	37	35.6	337	8	ADN46602	Adn46602 Thermococ	657	37	35.6	929	5	ABG30539	Abg30539 Alpha-iso
585	37	35.6	361	8	ADP45523	Adp45523 Human col	658	37	35.6	929	5	ABG30539	Abg30539 Alpha-iso
586	37	35.6	364	6	ADA34721	Ada34721 Acinetoba	659	37	35.6	965	5	ABG30565	Abg30565 Alpha-iso
587	37	35.6	366	9	ABD31464	Abd31464 Rubber al	660	37	35.6	965	5	ADA26483	Ada26483 Alpha-iso
588	37	35.6	374	5	ABP93667	Abp93667 Herbicida	661	37	35.6	990	8	ADN73117	Adn73117 Thale cre
589	37	35.6	382	5	ABG93031	Abg93031 S. cerevi	662	37	35.6	990	8	ADT56287	Adt56287 Plant pol
590	37	35.6	382	8	ADN18703	Adn18703 Bacterial	663	37	35.6	1152	6	ABM67949	Abm67949 Photorhab
591	37	35.6	387	2	AAW38078	Aaw38078 Sarcosine	664	37	35.6	1289	6	ABK52641	Abk52641 Protein s
592	37	35.6	387	9	ADX02873	Adx02873 Novel mod	665	37	35.6	1289	7	ADK63056	Adk63056 Disease t
593	37	35.6	387	9	ABE07361	Ab07361 Novel mod	666	37	35.6	1873	2	AAW73055	Aaw73055 Rabbit sk
594	37	35.6	389	2	AAW52658	Aaw52658 Arthrobac	667	37	35.6	1873	2	AAW18390	Aaw18390 Rabbit sk
595	37	35.6	389	2	AAW79150	Aaw79150 Arthrobac	668	37	35.6	1873	2	AAW37711	Aaw37711 Rabbit sk
596	37	35.6	399	2	AAW76735	Aaw76735 N-methyl	669	37	35.6	1873	3	AAW77544	Aay77544 Rabbit sk
597	37	35.6	399	2	AAW25149	Aaw25149 Modified	670	37	35.6	1873	9	ABE26777	Ab626777 Rabbit sk
598	37	35.6	399	2	AAW25150	Aaw25150 Modified	671	37	35.6	1925	6	AAE37000	Aae37000 Micromono
599	37	35.6	399	2	AAW25148	Aaw25148 Arthrobac	672	37	35.6	2108	10	AEF66135	Aef66135 Chick Ovo
600	37	35.6	399	2	AAW25151	Aaw25151 Modified	673	37	35.6	2108	10	AEF66132	Aef66132 Chick Ovo
601	37	35.6	399	2	AAW71461	Aaw71461 A modifie	674	37	35.6	4032	8	ADV99897	Adv99897 Nanchangm
602	37	35.6	399	2	AAW71462	Aaw71462 A modifie	675	37	35.6	6532	10	AEF57850	Aef57850 Polyketid
603	37	35.6	399	2	AAW53163	Aaw53163 Arthrobac	676	36.5	35.1	32	9	ABE21381	Aeb21381 Mouse ant
604	37	35.6	399	2	AAW53156	Aaw53156 Arthrobac	677	36.5	35.1	32	9	ABE31139	Aeb31139 Antibody
605	37	35.6	399	2	AAW53159	Aaw53159 Arthrobac	678	36.5	35.1	52	6	ABU56917	Abu56917 BONT/A Hc
606	37	35.6	399	2	AAW53157	Aaw53157 Arthrobac	679	36.5	35.1	53	6	ABU56848	Abu56848 BONT/A Hc
607	37	35.6	399	2	AAW53155	Aaw53155 Arthrobac	680	36.5	35.1	55	5	ABP09210	Abp09210 Human ORF

681	36.5	35.1	91	2	AAW95478	Mouse der	754	36.5	35.1	543	7	ADD12876	Adi12876
682	36.5	35.1	94	4	AAW88244	Human imm	755	36.5	35.1	662	4	AAE00477	Aae00477
683	36.5	35.1	107	4	AAW27353	Light cha	756	36.5	35.1	741	7	ABO61337	AbO61337
684	36.5	35.1	108	8	ADT74998	Light cha	757	36.5	35.1	777	7	ADC08297	AdC08297
685	36.5	35.1	108	8	ADT75001	Light cha	758	36.5	35.1	777	7	ADC07858	AdC07858
686	36.5	35.1	109	8	ADT75009	Light cha	759	36.5	35.1	1554	9	ADZ44908	AdZ44908
687	36.5	35.1	109	8	ADT74995	Light cha	760	36	34.6	16	10	AEF59557	Aef59557
688	36.5	35.1	109	8	ADT75003	Light cha	761	36	34.6	18	8	ABO54964	AbO54964
689	36.5	35.1	110	9	ADT75003	Light cha	762	36	34.6	18	10	AEQ00730	Aeg00730
690	36.5	35.1	110	9	ADY74788	Rat anti-	763	36	34.6	37	3	AAE16911	Aae16911
691	36.5	35.1	111	2	AAE65172	Murine NM	764	36	34.6	59	5	AAU83443	Aau83443
692	36.5	35.1	111	2	AAE75059	Light cha	765	36	34.6	59	5	AAU83440	Aau83440
693	36.5	35.1	112	8	AAE75059	Chimeric	766	36	34.6	59	5	AAU83441	Aau83441
694	36.5	35.1	112	2	AAW09156	Human IGE	767	36	34.6	63	4	AAU83442	Aau83442
695	36.5	35.1	112	2	AAW00830	Variable	768	36	34.6	63	4	AAU56121	Aau56121
696	36.5	35.1	112	2	AAW19016	Anti-huma	769	36	34.6	63	6	ABM52640	Abm52640
697	36.5	35.1	112	2	AAW27358	Light cha	770	36	34.6	80	5	ABM53994	Abm53994
698	36.5	35.1	112	8	ADR38689	Mouse lig	771	36	34.6	86	5	ABP01272	Abp01272
699	36.5	35.1	112	8	ADR38681	Mouse lig	772	36	34.6	94	4	AAW14840	Aaw14840
700	36.5	35.1	112	8	ADR38680	Mouse lig	773	36	34.6	94	4	ABB33808	Abb33808
701	36.5	35.1	112	8	ADR38679	Mouse lig	774	36	34.6	94	4	AAW27267	Aaw27267
702	36.5	35.1	112	8	ADT75130	Light cha	775	36	34.6	94	4	ABB28624	Abb28624
703	36.5	35.1	112	8	ADT75138	Light cha	776	36	34.6	94	4	ABB19250	Abb19250
704	36.5	35.1	112	8	ADT75123	Light cha	777	36	34.6	94	4	AAW66980	Aaw66980
705	36.5	35.1	112	8	ADT75139	Light cha	778	36	34.6	94	4	AAW54574	Aaw54574
706	36.5	35.1	112	8	ADT75137	Light cha	779	36	34.6	94	4	AAW02565	Aaw02565
707	36.5	35.1	112	8	ADT75136	Light cha	780	36	34.6	94	4	ABG48643	Abg48643
708	36.5	35.1	113	8	ADT75140	Light cha	781	36	34.6	94	5	ABG36637	Abg36637
709	36.5	35.1	113	8	ADT75124	Light cha	782	36	34.6	100	7	ADM05782	Adm05782
710	36.5	35.1	113	8	ADT75131	Light cha	783	36	34.6	100	9	AEC88712	Aec88712
711	36.5	35.1	113	8	ADT75129	Light cha	784	36	34.6	105	4	AAO07619	Aao07619
712	36.5	35.1	113	8	ADT75127	Light cha	785	36	34.6	107	4	AAW95420	Aaw95420
713	36.5	35.1	113	8	ADT75056	Light cha	786	36	34.6	113	9	ADX02189	Adx02189
714	36.5	35.1	113	8	ADT75135	Light cha	787	36	34.6	115	5	ABB49749	Abb49749
715	36.5	35.1	113	8	ADT75141	Light cha	788	36	34.6	117	8	ABO55085	AbO55085
716	36.5	35.1	113	9	ABE21354	Mouse ant	789	36	34.6	118	4	AAW94043	Aaw94043
717	36.5	35.1	113	9	ABE31112	Antibody	790	36	34.6	126	4	ABG11531	Abg11531
718	36.5	35.1	121	2	AAW33346	Sequence	791	36	34.6	126	7	ADC33270	AdC33270
719	36.5	35.1	121	2	AAW48616	Sequence	792	36	34.6	131	10	ABG01392	Abg01392
720	36.5	35.1	121	2	AAW48615	Sequence	793	36	34.6	143	7	ABO82352	AbO82352
721	36.5	35.1	121	2	AAW09259	Monoclon	794	36	34.6	156	3	AGW23166	Agw23166
722	36.5	35.1	121	2	AAW09274	MAB NM-01	795	36	34.6	157	10	AEF28744	Aef28744
723	36.5	35.1	132	4	AAO00490	Human pol	796	36	34.6	157	10	AEF28738	Aef28738
724	36.5	35.1	132	9	ADY74777	Rat anti-	797	36	34.6	157	10	AEF28736	Aef28736
725	36.5	35.1	133	7	ADH86121	Enterococ	798	36	34.6	157	10	AEF28739	Aef28739
726	36.5	35.1	217	9	ADY74780	Rat anti-	799	36	34.6	164	9	ABM91555	Abm91555
727	36.5	35.1	260	2	AAW77617	Anti-CS M	800	36	34.6	168	4	ABW79066	Abw79066
728	36.5	35.1	289	3	AAW43642	Human can	801	36	34.6	168	9	AEC56502	Aec56502
729	36.5	35.1	294	8	ADP81105	Protein o	802	36	34.6	171	4	AAW79737	Aaw79737
730	36.5	35.1	294	8	ADP81106	Protein o	803	36	34.6	171	4	AAW79759	Aaw79759
731	36.5	35.1	326	4	AAW68532	Human GTP	804	36	34.6	172	8	ADR88776	Adr88776
732	36.5	35.1	326	4	AAW93260	Human pro	805	36	34.6	175	3	ADR23165	Adr23165
733	36.5	35.1	331	4	ADM20095	Protein e	806	36	34.6	180	8	ADR95375	Adr95375
734	36.5	35.1	331	4	ADM19849	Protein e	807	36	34.6	180	9	AEA59245	Aea59245
735	36.5	35.1	361	6	ADU25118	Bacterial	808	36	34.6	183	3	AAW63109	Aaw63109
736	36.5	35.1	411	6	ABU17609	Protein e	809	36	34.6	189	6	ABU35100	Abu35100
737	36.5	35.1	411	6	ABU17923	Protein e	810	36	34.6	197	7	ADH87068	Adh87068
738	36.5	35.1	430	8	ADW88503	Streptococ	811	36	34.6	198	5	AAE18984	Aae18984
739	36.5	35.1	430	8	ADW81914	Streptococ	812	36	34.6	200	9	ABE40737	AbE40737
740	36.5	35.1	430	8	ADW79756	Streptococ	813	36	34.6	201	9	ABE37422	AbE37422
741	36.5	35.1	438	5	ABP26430	Streptococ	814	36	34.6	206	9	AEC95647	Aec95647
742	36.5	35.1	442	6	ADW24611	Bacterial	815	36	34.6	209	3	AAW53265	Aaw53265
743	36.5	35.1	444	6	ABM67311	Photorhab	816	36	34.6	209	3	AAW04568	Aaw04568
744	36.5	35.1	448	8	ADS14812	Pseudomon	817	36	34.6	216	8	ADS23713	Ads23713
745	36.5	35.1	462	6	ADA35841	Acinetoba	818	36	34.6	235	3	AAW53264	Aaw53264
746	36.5	35.1	464	6	ABU19266	Protein e	819	36	34.6	235	9	ABM96961	Abm96961
747	36.5	35.1	468	7	ABO68462	Pseudomon	820	36	34.6	236	3	AAW04567	Aaw04567
748	36.5	35.1	476	8	ADS42001	Bacterial	821	36	34.6	242	8	ADL04591	Adl04591
749	36.5	35.1	477	8	ADS22387	Bacterial	822	36	34.6	247	3	AAW19300	Aaw19300
750	36.5	35.1	478	5	ABP64914	Human pro	823	36	34.6	247	3	AAW36747	Aaw36747
751	36.5	35.1	480	8	ADS25724	Bacterial	824	36	34.6	247	3	AAW40096	Aaw40096
752	36.5	35.1	483	8	ADS26203	Bacterial	825	36	34.6	249	2	AAW5065	Aaw5065
753	36.5	35.1	489	7	ABO68163	Pseudomon	826	36	34.6	249	5	ABP54559	Abp54559

827	36	34.6	249	7	ADC45087	S. pneumo	Adc45087	900	36	34.6	437	5	AAO20319	Protein o
828	36	34.6	250	2	AAW61251	Streptoco	Aaw61251	901	36	34.6	437	5	ABG80323	C. glutam
829	36	34.6	250	5	ABP54669	S. pneumo	Abp54669	902	36	34.6	437	8	ADR28244	Corynebac
830	36	34.6	250	7	ADC45307	S. pneumo	Adc45307	903	36	34.6	437	9	AEC56498	Corynebac
831	36	34.6	250	7	ABO84368	Pseudomon	Ab084368	904	36	34.6	447	4	AEC53827	Human pol
832	36	34.6	260	6	ADX02026	SARS coro	Adx02026	905	36	34.6	447	5	AAE21720	Human pol
833	36	34.6	266	8	ADK48784	Streptoco	Adk48784	906	36	34.6	447	7	ADM05915	Human pro
834	36	34.6	267	3	RAY81677	Streptoco	Ray81677	907	36	34.6	447	9	AEC88845	Human CDN
835	36	34.6	268	8	ABO58929	Human gen	Ab058929	908	36	34.6	451	3	AAG39206	Arabidops
836	36	34.6	270	5	AAU76160	Bacillus	Aau76160	909	36	34.6	453	7	ABO71768	Pseudomon
837	36	34.6	271	6	ABU00996	S. pneumo	Abu00996	910	36	34.6	453	8	ADN27283	Bacterial
838	36	34.6	271	8	ADN23178	Bacterial	Adn23178	911	36	34.6	454	8	ADS44648	Bacterial
839	36	34.6	280	2	AAE37292	Barley ty	Aae37292	912	36	34.6	456	6	ABP57494	Mycobacte
840	36	34.6	280	2	AAE63904	Type I ri	Aae63904	913	36	34.6	456	6	AAU16378	Murine pa
841	36	34.6	280	2	AAE74178	Barley Ty	Aae74178	914	36	34.6	457	6	ABU96704	Human nuc
842	36	34.6	280	2	AAW21698	Barley tr	Aaw21698	915	36	34.6	460	2	AAE69874	B.thermog
843	36	34.6	280	2	AAW25135	Barley Tr	Aaw25135	916	36	34.6	460	4	AAW40613	Human pol
844	36	34.6	281	2	AAE25048	Protein s	Aae25048	917	36	34.6	460	6	AAO16374	Human pan
845	36	34.6	281	2	AAE52575	Protein s	Aae52575	918	36	34.6	466	3	AAE29656	Arabidops
846	36	34.6	292	6	ADA35355	Acinetoba	Ada35355	919	36	34.6	468	4	ABB61357	Drosophil
847	36	34.6	295	4	ABU65260	Novel hum	Abu65260	920	36	34.6	470	8	ADS27421	Bacterial
848	36	34.6	297	6	ABE65990	Drosophil	Ab65990	921	36	34.6	471	6	ABU45496	Protein e
849	36	34.6	300	8	ADX72798	Plant ful	Adx72798	922	36	34.6	472	3	AAU78103	Bacillus
850	36	34.6	302	10	ABE97921	Mesorhizo	Abe97921	923	36	34.6	477	3	AAE42562	Arabidops
851	36	34.6	314	3	ABE53189	Macaca mu	Ab53189	924	36	34.6	479	5	ABP27212	Streptoco
852	36	34.6	320	3	ABE97327	Novel hum	Ab97327	925	36	34.6	479	8	ADV82130	Streptoco
853	36	34.6	320	3	AAE36746	Arabidops	Aag36746	926	36	34.6	483	2	AAW31538	Amino aci
854	36	34.6	320	3	AAE19299	Arabidops	Aag19299	927	36	34.6	483	6	ABG74403	Banana 13
855	36	34.6	320	8	ADN72599	Thale cre	Adn72599	928	36	34.6	484	9	ABW92032	M. xanthu
856	36	34.6	321	3	AAE40095	Arabidops	Aag40095	929	36	34.6	492	8	ADN18407	Bacterial
857	36	34.6	324	8	ADX74837	Plant ful	Adx74837	930	36	34.6	493	6	ABU33590	Protein e
858	36	34.6	325	3	AAE40094	Arabidops	Aag40094	931	36	34.6	493	9	ABE41691	L. pneumo
859	36	34.6	328	8	ADT60244	Plant pol	Adt60244	932	36	34.6	497	9	ABE38427	L. pneumo
860	36	34.6	328	6	ABU30270	Protein e	Abu30270	933	36	34.6	504	8	ADN20868	Bacterial
861	36	34.6	336	6	ABU23886	Protein e	Abu23886	934	36	34.6	505	8	ADV88738	Streptoco
862	36	34.6	354	4	AAE82604	Cattle Ig	Ab82604	935	36	34.6	505	8	ADV79991	Streptoco
863	36	34.6	354	4	AAE82606	Sheep Ig	Ab82606	936	36	34.6	506	4	AAE78226	Lepomis c
864	36	34.6	357	2	ABO04127	Fusarium	Ab004127	937	36	34.6	507	3	AAE29655	Arabidops
865	36	34.6	360	3	AAE42563	Arabidops	Aag42563	938	36	34.6	519	7	AAJ68604	Human hea
866	36	34.6	369	7	ABO81335	Pseudomon	Ab081335	939	36	34.6	519	8	ADQ37886	Variant h
867	36	34.6	371	6	ABU79127	Lipid A b	Abu79127	940	36	34.6	519	8	ADQ37881	Human DHP
868	36	34.6	371	7	ADF43387	Lipid A b	Adf43387	941	36	34.6	531	4	ABB63988	Drosophil
869	36	34.6	371	8	ADT49844	Murine TA	Adt49844	942	36	34.6	539	3	AAE29654	Arabidops
870	36	34.6	371	9	AEA03058	Lipid A b	Aea03058	943	36	34.6	540	2	AAE72678	Protein d
871	36	34.6	373	2	AAW88213	Pseudomon	Aaw88213	944	36	34.6	549	4	ABE71680	Drosophil
872	36	34.6	374	7	ADC01328	Enterobac	Adc01328	945	36	34.6	556	7	ABE80720	Microeate
873	36	34.6	374	8	ADH88970	Escherich	Adh88970	946	36	34.6	557	2	AAW57844	F24A muta
874	36	34.6	376	2	AAE15272	Fusarium	Aae15272	947	36	34.6	557	2	AAW57840	Wild type
875	36	34.6	376	2	AAE25429	Cellulase	Aae25429	948	36	34.6	557	2	AAW57846	F24L muta
876	36	34.6	376	2	AAE25527	Fusarium	Aae25527	949	36	34.6	557	2	AAW57845	F24V muta
877	36	34.6	376	2	AAE25466	Endogluc	Aae25466	950	36	34.6	557	2	AAW57843	V56T muta
878	36	34.6	376	2	AAE27969	Endogluc	Aae27969	951	36	34.6	557	7	ADA38225	Escherich
879	36	34.6	376	2	AAE24064	Endogluc	Aae24064	952	36	34.6	566	8	ADS23484	Bacterial
880	36	34.6	376	2	AAE37151	Dye trans	Aae37151	953	36	34.6	570	4	ABB63990	Drosophil
881	36	34.6	376	2	AAE67389	F. oxyspo	Aay67389	954	36	34.6	571	4	ABG00360	Novel hum
882	36	34.6	376	2	AAW46617	Fusarium	Aaw46617	955	36	34.6	582	4	AAE78227	Lepomis c
883	36	34.6	378	8	ADT49883	Murine TA	Adt49883	956	36	34.6	588	9	ADX06294	Cyclin-de
884	36	34.6	401	7	ABO84231	Pseudomon	Ab084231	957	36	34.6	599	4	ABE71947	Drosophil
885	36	34.6	406	9	AEC95648	Calcium c	Aec95648	958	36	34.6	599	9	AEA00258	Leishmani
886	36	34.6	409	2	AAW01501	50 kD end	Aaw01501	959	36	34.6	603	5	ADL23399	Plant AMP
887	36	34.6	409	2	AAW17927	Abbrading	Aaw17927	960	36	34.6	603	7	ADL72426	Arabidops
888	36	34.6	411	2	AAE83399	Fusarium	Aae83399	961	36	34.6	603	7	ADL72277	Arabidops
889	36	34.6	420	8	ADS21028	Bacterial	Ad21028	962	36	34.6	605	6	ABM04806	Rat ribop
890	36	34.6	423	6	ABP80988	N. gonorr	Abp80988	963	36	34.6	605	7	ADE63298	Rat Prote
891	36	34.6	424	6	ABU24236	Protein e	Abu24236	964	36	34.6	605	8	ADT57496	Plant pol
892	36	34.6	425	4	ABG29201	Novel hum	Abg29201	965	36	34.6	607	8	ADM57739	Human cal
893	36	34.6	425	8	ADS30686	Bacterial	Ad30686	966	36	34.6	608	9	ADM25843	Hybrid al
894	36	34.6	426	6	ABU19824	Protein'e	Abu19824	967	36	34.6	608	9	ADM25847	Hybrid al
895	36	34.6	427	2	AAU15239	Fusarium	Aau15239	968	36	34.6	608	9	ADM25846	Hybrid al
896	36	34.6	433	2	AAW12400	A. chryso	Aaw12400	969	36	34.6	608	9	ADM25851	Hybrid al
897	36	34.6	437	4	AAE79064	Corynebac	Aae79064	970	36	34.6	608	9	ADM25851	Hybrid al
898	36	34.6	437	4	AAU71863	C. glutam	Aau71863	971	36	34.6	620	4	AAW50111	Feline pa
899	36	34.6	437	4	AAE90472	C glutami	Aag90472	972	36	34.6	628	4	AAU37023	Staphyloc





KW cytotatic; antiinflammatory; vulnery; antiarteriosclerotic;  
KW vasotrophic; immunosuppressive; nephrotropic; antidiabetic; human.  
OS Homo sapiens.

XX US2004213789-A1.

XX 28-OCT-2004.

XX 22-AUG-2003; 2003US-00645659.

XX 02-SEP-1997; 97US-00922170.

XX 01-MAY-1998; 98US-00071739.

XX 04-NOV-1998; 98US-00186200.

XX 19-FEB-2003; 2003US-00368044.

XX (YACO) YACOB-ZEEVI O.

PA (PERE) PERETZ T.

PA (MIRO) MIRON D.

PA (SHLO) SHLOMI Y.

PA (PECK) PECKER I.

PA (AYAL) AYAL-HERSHKOVITZ M.

PA (FEIN) FEINSTEIN E.

PA (GELD) GELDER J M V.

PA (VLAD) VLADAVSKY I.

PA (FRIE) FRIEDMANN Y.

XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;

PI Ayal-Hershkovitz M, Feinstein E, Gelder JMV, Vlodayvsky I;

PI Friedmann Y;

DR WPI; 2004-774790/76.

XX New neutralizing monoclonal anti-heparanase antibodies, useful for

PT detecting, treating or preventing cancer, inflammatory or autoimmune

PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.

XX Claim 53; SEQ ID NO 7; 68pp; English.

XX The invention relates to an isolated antibody or antibody portion capable

CC of specifically binding to or elicited by at least one epitope of a

CC heparanase protein, where the heparanase protein is at least 60%

CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and

CC where at least one epitope comprises a sequence at least 70% homologous

CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)

CC a hybridoma cell line comprising a cell line for producing the monoclonal

CC antibody, (b) a method for detecting, treating or preventing a

CC pathological condition or a heparanase-related disorder or condition in a

CC subject, (c) a method for monitoring the state of a heparanase-related

CC disorder or condition in a subject, and (d) a pharmaceutical composition

CC comprising the isolated anti-heparanase antibody or antibody portion and

CC a pharmaceutical carrier. The antibody, methods, and composition are

CC useful for detecting, treating, preventing or monitoring a pathological

CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition

CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,

CC or prostate cancer), minor cell proliferation, invasion of circulating

CC tumour cells, or a metastatic disease, or a heparanase-related disorder

CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy

XX (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or

XX renal disease or disorder (diabetic nephropathy, glomerulosclerosis,

XX nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell

XX carcinoma) in a mammal. This sequence represents a functional peptide

XX epitope of human heparanase.

XX Sequence 19 AA;

Query Match

Best Local Similarity 100.0%; Score 104; DB 8; Length 19;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFLIFDPK 19

|||||

Db 1 PAYLRFGGTKTDFLIFDPK 19

## RESULT 3

AEA42429

ID AEA42429 standard; peptide; 19 AA.

XX AEA42429;

XX 28-JUL-2005 (first entry)

XX Human heparanase epitope peptide SEQ ID NO:7.

XX antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;  
KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotrophic;  
KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;  
KW angiogenesis disorder; cancer; tumor; metastasis; epitope.

XX Homo sapiens.

XX AU2004201462-A1.

XX 06-MAY-2004.

XX 08-APR-2004; 2004AU-00201462.

XX 08-APR-2004; 2004AU-00201462.

XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

PA (HADA-) HADASIT MEDICAL RES SERVICES &amp; DEV LTD.

XX Vlodayvsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;  
PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;  
PI Feinstein E;

XX WPI; 2005-173343/19.

XX Novel isolated antibody capable of specifically binding to epitope of  
PT heparanase protein, useful for preventing and treating heparanase-related  
PT disorder such as inflammatory disorder, scars, autoimmune conditions or  
PT angiogenesis.

XX Claim 7; SEQ ID NO 7; 260pp; English.

XX The invention relates to an isolated antibody or its portion (I) capable  
CC of specifically binding to an epitope of a heparanase protein. Also  
CC described: (1) a cell line (II) for producing a monoclonal antibody or  
CC its portion, comprising a cell line for producing (I); (2) a

CC pharmaceutical composition comprising (I) and a carrier; and (3) an

CC affinity medium (III) for binding human heparanase polypeptides,

CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)

CC useful for treating a subject suffering from a pathological condition,

CC which involves administering (I) to the subject. (I) is useful for

CC preventing and treating heparanase-related disorder or condition chosen

CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune

CC condition, angiogenesis, cell proliferation, cancerous condition, tumor

CC cell proliferation, invasion of circulating tumor cells and metastatic

CC disease. (I) is useful for detecting the presence of heparanase

CC polypeptide in a sample. (I) is useful for detecting heparanase-related

CC disease or condition in a subject such as vertebrate, preferably mammal

CC e.g., human. The heparanase-related disorder or condition further

CC includes renal disease or disorder chosen from diabetic nephropathy,

CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome

CC and renal cell carcinoma. The present sequence represents a human

CC heparanase epitope peptide, which is used in the exemplification of the

XX present invention.

SQ Sequence 19 AA;

Query Match

Best Local Similarity 100.0%; Score 104; DB 9; Length 19;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFLIFDPK 19

Db  
RESULT 4  
ADR88217  
ID ADR88217 standard; protein; 74 AA.  
XX  
AC  
XX  
AD88217;  
XX  
DT  
XX  
18-NOV-2004 (first entry)  
XX  
DE  
XX  
Human mature processed heparanase dimer 8 kDa subunit.

XX  
XX  
Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;  
XX  
autoimmune disorder; cancer; angiogenesis; metastatic disease;  
XX  
atherosclerosis; testonosis; aneurysm; solid cancer; non-solid cancer;  
XX  
haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;  
XX  
Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;  
XX  
human; heparanase; enzyme.

XX  
Homo sapiens.

XX  
US2004170631-A1.

XX  
02-SEP-2004.

XX  
28-NOV-2003; 2003US-00722502.

XX  
02-SEP-1997; 97US-00922170.

XX  
01-MAY-1998; 98US-00071739.

XX  
04-NOV-1998; 98US-00186200.

XX  
19-FEB-2003; 2003US-00368044.

XX  
22-AUG-2003; 2003US-00645659.

XX  
(YACO/) YACOBY-ZEEVI O.

XX  
(PERE/) PERETZ T.

XX  
(MIRO/) MIRON D.

XX  
(SHLO/) SHLOMI Y.

XX  
(PECK/) PECKER I.

XX  
(AYAL/) AYAL-HERSHKOVITZ M.

XX  
(FEIN/) FEINSTEIN E.

XX  
(VGLD/) VAN GELDER J M.

XX  
(VLOD/) VLODAVSKY I.

XX  
(FRIE/) FRIEDMANN Y.

XX  
Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;

PI  
Ayal-Herskovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;

PI  
Friedmann Y;

XX  
WPI; 2004-625084/60.

XX  
Targeted drug delivery to a heparanase-expressing tissue of a patient,

PT  
useful for treating heparanase-associated conditions such as inflammation

PT  
or cancer, comprises administering a drug and an anti-heparanase antibody

PT  
complex.

XX  
Claim 2; SEQ ID NO 11; 58pp; English.

XX  
The invention relates to a method of targeted drug delivery to a tissue

XX  
of a patient, the tissue expressing heparanase. The method comprises

CC  
providing a complex of a drug directly or indirectly linked to an anti-

CC  
heparanase antibody, and administering the complex to the patient. In the

CC  
targeted drug delivery, the antibody comprises an antibody or its portion

CC  
capable of specifically binding to at least one epitope of a heparanase

CC  
protein. The composition and methods of the invention are useful for

CC  
diagnosing, preventing or treating conditions associated with heparanase

CC  
catalytic activity (e.g. an inflammatory disorder, wound, scar,

CC  
vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell

CC  
proliferation, invasion of circulating tumour cells and metastatic

CC  
disease), for purifying heparanase, or for developing drugs for those

CC  
heparanase-associated conditions. The vasculopathy is atherosclerosis,

CC  
restenosis or aneurysm. The cancerous condition is a solid cancer or a

CC  
restenosis or aneurysm.

CC  
non-solid cancer. The non-solid cancer is a haematopoietic malignancy  
CC  
selected from acute lymphocytic leukaemia (ALL), acute myelogenous  
CC  
leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous  
CC  
leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,  
CC  
Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and  
CC  
multiple myeloma. The solid cancer is selected from tumours in lip and  
CC  
oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,  
CC  
thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,  
CC  
anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of  
CC  
Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue  
CC  
sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,  
CC  
vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic  
CC  
tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary  
CC  
bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,  
CC  
malignant melanoma of the conjunctiva, malignant melanoma of the uvea,  
CC  
retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,  
CC  
brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's  
CC  
sarcoma. The present sequence is the 8 kDa subunit of human mature  
CC  
processed heparanase dimer.

XX  
Sequence 74 AA;

Query Match 100.0%; Score 104; DB 8; Length 74;

Best Local Similarity 100.0%; Pred. No. 7.4e-10;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKDTFLIFDPK 19

Db 54 PAYLRFGGTKDTFLIFDPK 72

RESULT 5

ADT78184

ID ADT78184 standard; protein; 74 AA.

XX  
AC ADT78184;

XX  
DT 13-JAN-2005 (first entry)

XX  
DE 8kDa subunit of mature processed human heparanase dimer.

XX  
KW Antibody; epitope; heparanase; pathological condition; angiogenesis;  
XX  
cell proliferation; cancerous condition; tumour cell invasion;  
XX  
metastatic disease; heparanase-related disorder; inflammatory disorder;  
XX  
wound; scar; vasculopathy; autoimmune condition; renal disease;  
XX  
cystostatic; antiinflammatory; vulnary; antiarteriosclerotic;  
XX  
vasotrophic; immunosuppressive; nephrotropic; antidiabetic; human.

XX  
OS Homo sapiens.

XX  
PN US2004213789-A1.

XX  
PD 28-OCT-2004.

XX  
PF 22-AUG-2003; 2003US-00645659.

XX  
PR 02-SEP-1997; 97US-00922170.

XX  
PR 01-MAY-1998; 98US-00071739.

XX  
PR 04-NOV-1998; 98US-00186200.

XX  
PR 19-FEB-2003; 2003US-00368044.

XX  
(YACO/) YACOBY-ZEEVI O.

PA (PERE/) PERETZ T.

PA (MIRO/) MIRON D.

PA (SHLO/) SHLOMI Y.

PA (PECK/) PECKER I.

PA (AYAL/) AYAL-HERSHKOVITZ M.

PA (FEIN/) FEINSTEIN E.

PA (GELD/) GELDER J M V.

PA (VLOD/) VLODAVSKY I.

XX  
(FRIE/) FRIEDMANN Y.

PI  
Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;

PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;  
PI Friedmann Y;  
XX WPI; 2004-774790/76.  
DR  
XX  
XX New neutralizing monoclonal anti-heparanase antibodies, useful for  
PT detecting, treating or preventing cancer, inflammatory or autoimmune  
PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.  
XX  
PS Claim 5; SEQ ID NO 11; 68pp; English.  
XX  
XX The invention relates to an isolated antibody or antibody portion capable  
CC of specifically binding to or elicited by at least one epitope of a  
CC heparanase protein, where the heparanase protein is at least 60%  
CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and  
CC where at least one epitope comprises a sequence at least 70% homologous  
CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)  
CC a hybridoma cell line comprising a cell line for producing the monoclonal  
CC antibody, (b) a method for detecting, treating or preventing a  
CC pathological condition or a heparanase-related disorder or condition in a  
CC subject, (c) a method for monitoring the state of a heparanase-related  
CC disorder or condition in a subject, and (d) a pharmaceutical composition  
CC comprising the isolated anti-heparanase antibody or antibody portion and  
CC a pharmaceutical carrier. The antibody, methods, and composition are  
CC useful for detecting, treating, preventing or monitoring a pathological  
CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition  
CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,  
CC or prostate cancer), minor cell proliferation, invasion of circulating  
CC tumour cells, or a metastatic disease, or a heparanase-related disorder  
CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy  
CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or  
CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,  
CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell  
CC carcinoma) in a mammal. This sequence represents the 8kDa subunit of  
CC mature processed human heparanase dimer.  
XX  
SQ Sequence 74 AA;  
  
Query Match 100.0%; Score 104; DB 8; Length 74;  
Best Local Similarity 100.0%; Pred. No. 7.4e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PAYLRFGGTKTDFLIFDPK 19  
Db |||||  
54 PAYLRFGGTKTDFLIFDPK 72  
  
RESULT 6  
ADY27059  
ID ADY27059 standard; protein; 74 AA.  
XX  
XX AC ADY27059;  
XX  
DT 05-MAY-2005 (first entry)  
XX  
XX Heparanase inhibitor protein #2.  
XX  
XX Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
KW neurological disease; viral infection; infection; cytostatic;  
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
KW heparanase modulator; enzyme purification.  
XX  
XX OS Homo sapiens.  
XX  
XX WO2005016227-A2.  
PN  
XX  
PD 24-FEB-2005.  
XX  
XX 12-AUG-2004; 2004WO-IL000744.  
PF  
XX  
PR 14-AUG-2003; 2003US-0494800P.  
PR  
PR 12-JAN-2004; 2004US-0535492P.  
XX

(INSI-) INSIGHT BIOPHARMACEUTICALS LTD.  
XX  
PI Van-Gelder JM, Miron D;  
XX  
XX WPI; 2005-182203/19.  
DR  
XX  
XX Regulating heparanase activity, useful for treating heparanase-associated  
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
PT neurological diseases or viral diseases) comprises modulating heparanase  
PT activation.  
XX  
PS Claim 55; SEQ ID NO 35; 211pp; English.  
XX  
XX The invention relates to a method of regulating heparanase activity in a  
CC tissue or regulating a biological process depending at least in part on  
CC heparanase activity comprising modulating heparanase activation. The  
CC invention also relates to methods of treating a heparanase- or heparin  
CC binding protein-associated disease or disorder in a subject, a  
CC pharmaceutical composition for use in the treatment of a heparanase-  
CC associated disease or disorder comprising a therapeutic amount of an  
CC agent capable of modulating heparanase activation and a pharmaceutical  
CC carrier or diluent, a method of identifying a protease activator of  
CC heparanase, a protease substrate mimetic comprising a peptide  
CC representing a subset or all substrate residues or cleavage sites of  
CC human heparanase or an equivalent non-human heparanase, a method of  
CC producing active heparanase and a method of modulating an adhesion  
CC activity of heparanase. The composition and methods are useful for  
CC modulating heparanase activation and for treating heparanase-associated  
CC diseases or disorders such as cancer, inflammation, cardiovascular  
CC diseases, neurological diseases or viral infections. This sequence  
CC represents a heparanase inhibitor protein used in the scope of the  
CC invention.  
XX  
SQ Sequence 74 AA;  
  
Query Match 100.0%; Score 104; DB 9; Length 74;  
Best Local Similarity 100.0%; Pred. No. 7.4e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PAYLRFGGTKTDFLIFDPK 19  
Db |||||  
54 PAYLRFGGTKTDFLIFDPK 72  
  
RESULT 7  
ADZ18994  
ID ADZ18994 standard; protein; 74 AA.  
XX  
XX AC ADZ18994;  
XX  
DT 16-JUN-2005 (first entry)  
XX  
XX Human heparanase consensus cleavage site #1.  
XX  
XX Enzyme engineering; heparanase; metastasis; autoimmune disease;  
KW inflammation; neoplasm; immune disorder; antinflammatory; cytostatic;  
KW immunosuppressive; enzyme.  
XX  
XX OS Homo sapiens.  
XX  
XX WO2005030962-A1.  
PN  
XX  
PD 07-APR-2005.  
XX  
XX 17-SEP-2004; 2004WO-EP010517.  
PF  
XX  
XX 26-SEP-2003; 2003US-0506479P.  
PR  
PR 20-JAN-2004; 2004US-0537729P.  
XX  
XX (RICE-) IST RICERCHÉ BIOL MOLECOLARE ANGELETTI.  
PA  
XX Lahm A, Nardella C, Pallaro M, Steinkuhler C;  
PI  
XX

DR WPI; 2005-273382/28.  
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
PT nucleotide sequence that encodes mammalian heparanase protein and has two  
PT consensus cleavage sites located between specific nucleotide encoding  
PT residues.  
XX Disclosure; SEQ ID NO 15; 65pp; English.  
XX The invention relates to a synthetic nucleic acid molecule that encodes  
CC mammalian heparanase protein, where the nucleic acid comprises two  
CC consensus cleavage sites recognized by endoproteinase. The sequences are  
CC useful for expressing mammalian heparanase in non-mammalian cells and in  
CC inhibitor screening assays for the development of therapeutics or  
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
CC and/or inflammation. This sequence represents a human heparanase  
CC consensus cleavage site used in the scope of the invention.  
XX Sequence 74 AA;  
SQ  
Query Match 100.0%; Score 104; DB 9; Length 74;  
Best Local Similarity 100.0%; Pred. No. 7.4e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PAYLRFGGKTDFLIIDPK 19  
Db 54 PAYLRFGGKTDFLIIDPK 72  
RESULT 8  
ID AEA42433  
AA AEA42433 standard; protein; 74 AA.  
XX AEA42433;  
AC  
XX  
DT 28-JUL-2005 (first entry)  
XX Human mature heparanase dimer 8 kDa subunit SEQ ID NO:11.  
DE  
XX antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;  
KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;  
KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;  
KW angiogenesis disorder; cancer; tumor; metastasis.  
XX  
XX Homo sapiens.  
OS  
XX AU2004201462-A1.  
PN  
XX 06-MAY-2004.  
PD  
XX 08-APR-2004; 2004AU-00201462.  
PF  
XX 08-APR-2004; 2004AU-00201462.  
PR  
XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.  
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.  
XX Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;  
PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;  
PI Feinstein E;  
XX WPI; 2005-173343/19.  
DR  
XX Novel isolated antibody capable of specifically binding to epitope of  
PT heparanase protein, useful for preventing and treating heparanase-related  
PT disorder such as inflammatory disorder, scars, autoimmune conditions or  
PT angiogenesis.  
XX  
XX Claim 2; SEQ ID NO 11; 260pp; English.  
PS  
XX The invention relates to an isolated antibody or its portion (I) capable  
CC of specifically binding to an epitope of a heparanase protein. Also  
CC described: (1) a cell line (II) for producing a monoclonal antibody or

CC its portion, comprising a cell line for producing (I); (2) a  
CC pharmaceutical composition comprising (I) and a carrier; and (3) an  
CC affinity medium (III) for binding human heparanase polypeptides,  
CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)  
CC useful for treating a subject suffering from a pathological condition,  
CC which involves administering (I) to the subject. (I) is useful for  
CC preventing and treating heparanase-related disorder or condition chosen  
CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune  
CC condition, angiogenesis, cell proliferation, cancerous condition, tumor  
CC cell proliferation, invasion of circulating tumor cells and metastatic  
CC disease. (II) is useful for detecting the presence of heparanase  
CC polypeptide in a sample. (I) is useful for detecting heparanase-related  
CC disease or condition in a subject such as vertebrate, preferably mammal  
CC e.g., human. The heparanase-related disorder or condition further  
CC includes renal disease or disorder chosen from diabetic nephropathy,  
CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome  
CC and renal cell carcinoma. The present sequence represents the 8 kDa  
CC subunit of the human mature processed heparanase dimer, which is used in  
CC the exemplification of the present invention.  
XX  
SQ Sequence 74 AA;  
Query Match 100.0%; Score 104; DB 9; Length 74;  
Best Local Similarity 100.0%; Pred. No. 7.4e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PAYLRFGGKTDFLIIDPK 19  
Db 54 PAYLRFGGKTDFLIIDPK 72  
RESULT 9  
ID ADZ19012 standard; protein; 174 AA.  
ADZ19012  
XX  
AC ADZ19012;  
XX  
DT 16-JUN-2005 (first entry)  
XX Human heparanase protein.  
DE  
XX Enzyme engineering; heparanase; metastasis; autoimmune disease;  
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
KW immunosuppressive; enzyme.  
XX  
XX Homo sapiens.  
OS  
XX WO2005030962-A1.  
PN  
XX 07-APR-2005.  
PD  
XX 17-SEP-2004; 2004WO-EP010517.  
PF  
XX 26-SEP-2003; 2003US-0506479P.  
PR 20-JAN-2004; 2004US-0537729P.  
XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.  
XX Lahm A, Nardella C, Pallaoro M, Steinkuhler C;  
PI WPI; 2005-273382/28.  
XX  
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
PT nucleotide sequence that encodes mammalian heparanase protein and has two  
PT consensus cleavage sites located between specific nucleotide encoding  
PT residues.  
XX Disclosure; SEQ ID NO 33; 65pp; English.  
PS  
XX The invention relates to a synthetic nucleic acid molecule that encodes  
CC mammalian heparanase protein, where the nucleic acid comprises two  
CC consensus cleavage sites recognized by endoproteinase. The sequences are  
CC useful for expressing mammalian heparanase in non-mammalian cells and in  
CC inhibitor screening assays for the development of therapeutics or  
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
CC and/or inflammation. This sequence represents a human heparanase  
CC consensus cleavage site used in the scope of the invention.  
XX Sequence 74 AA;  
SQ

CC inhibitor screening assays for the development of therapeutics or  
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
CC and/or inflammation. This sequence represents a human heparanase protein  
CC used in the scope of the invention.

XX SQ Sequence 174 AA;  
Query Match 100.0%; Score 104; DB 9; Length 174;  
Best Local Similarity 100.0%; Pred. No. 1.9e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19  
|||||  
Db 55 PAYLRFGGTKTDFLIFDPK 73

RESULT 10  
ADY27061  
ID ADY27061 standard; protein; 460 AA.

XX AC ADY27061;  
XX DT 05-MAY-2005 (first entry)  
XX DE Heparanase inhibitor protein #4.  
XX KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
KW neurological disease; viral infection; infection; cytostatic;  
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
KW heparanase modulator; enzyme purification.

XX OS Homo sapiens.

XX PN WO2005016227-A2.

XX PD 24-FEB-2005.

XX PF 12-AUG-2004; 2004WO-IL0000744.

XX PR 14-AUG-2003; 2003US-0494800P.

XX PR 12-JAN-2004; 2004US-0535492P.

XX PA (NSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX PI Van-Gelder JM, Miron D;

XX DR WPI; 2005-182203/19.

XX PT Regulating heparanase activity, useful for treating heparanase-associated  
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
PT neurological diseases or viral diseases) comprises modulating heparanase  
PT activation.

XX PS Disclosure; SEQ ID NO 37; 211pp; English.

XX CC The invention relates to a method of regulating heparanase activity in a  
CC tissue or regulating a biological process depending at least in part on  
CC heparanase activity comprising modulating heparanase activation. The  
CC invention also relates to methods of treating a heparanase- or heparin  
CC binding protein-associated disease or disorder in a subject, a  
CC pharmaceutical composition for use in the treatment of a heparanase-  
CC associated disease or disorder comprising a therapeutic amount of an  
CC agent capable of modulating heparanase activation and a pharmaceutical  
CC carrier or diluent, a method of identifying a protease activator of  
CC heparanase, a protease substrate mimetic comprising a peptide  
CC representing a subset or all substrate residues or cleavage sites of  
CC human heparanase or an equivalent non-human heparanase, a method of  
CC producing active heparanase and a method of modulating an adhesion  
CC activity of heparanase. The composition and methods are useful for  
CC modulating heparanase activation and for treating heparanase-associated  
CC diseases or disorders such as cancer, inflammation, cardiovascular  
CC diseases, neurological diseases or viral infections. This sequence  
CC represents a heparanase inhibitor protein used in the scope of the

CC invention.

XX SQ Sequence 460 AA;

Query Match 100.0%; Score 104; DB 9; Length 460;  
Best Local Similarity 100.0%; Pred. No. 5.8e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19  
|||||  
Db 54 PAYLRFGGTKTDFLIFDPK 72

RESULT 11

AEB87589

ID AEB87589 standard; protein; 486 AA.

XX AC AEB87589;

XX DT 06-OCT-2005 (first entry)

XX DE Human heparanase 65delta20 deletion mutant.

XX KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;  
KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;  
KW leukemia; lymphoma; dermatological disease; hematological disease;  
KW immune disorder; inflammation; antiinflammatory; renal disease;  
KW nephrotropic; endocrine disease; genitourinary disease;  
KW autoimmune disease; immunosuppressive; drug screening; mutein.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO2005071070-A2.

XX PD 04-AUG-2005.

XX PF 20-JAN-2005; 2005WO-IL0000068.

XX PR 22-JAN-2004; 2004IL-00160025.

XX PR 28-JUL-2004; 2004US-00901943.

XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX PI Vlodavsky I, Ilan N, Levy-Adam F;

XX DR WPI; 2005-564219/57.

XX DR N-PSDB; AEB87588.

XX PT New amino acid sequences derived from the 50 kDa subunit of heparanase,  
PT for treating or inhibiting malignant proliferative, inflammatory, kidney  
PT disorder or autoimmune disorder.

XX PS Claim 107; SEQ ID NO 31; 167pp; English.

XX CC The present sequence is that of a deletion mutant of human heparanase,  
CC denoted 65delta20, which is devoid of amino acid residues 411-432 of the  
CC native protein. The recombinant protein is deficient of heparanase  
CC endoglycosidase catalytic activity. The invention relates to amino acid  
CC sequences derived from the N-terminus region of the 50 kDa subunit of  
CC heparanase, particularly in the regions between amino acid residues 158-  
CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise  
CC heparin-binding domains. The invention also provides an antibody directed  
CC to these sequences, in particular the 158-171 peptide, and compositions  
CC and uses of this antibody as a heparanase inhibitor. A screening method  
CC is provided for specific heparanase inhibitors. Claimed pharmaceutical  
CC compositions comprising (i) a peptide derived from the N-terminus region  
CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a  
CC peptide, or (iii) an antibody which specifically recognizes the peptide  
CC are used for the inhibition of heparanase catalytic activity associated  
CC with an inflammatory disorder, kidney disease, autoimmune disease,  
CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or  
CC with a malignant proliferative disorder, especially a solid or non-solid

```

CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.
XX
SQ Sequence 486 AA;

    Query Match      100.0%; Score 104; DB 9; Length 486;
    Best Local Similarity 100.0%; Pred. No. 6.1e-09;
    Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIIFDPK 19
   |||||
Db 54 PAYLRFGGTKTDFLIIFDPK 72

RESULT 12
ADZ18996
ID ADZ18996 standard; protein; 492 AA.
XX
AC ADZ18996;
XX
DT 16-JUN-2005 (first entry)
XX
DE Hep106 construct protein.
XX
KW Enzyme engineering; heparanase; hep106; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX
OS Synthetic.
XX
PN WO2005030962-A1.
XX
PD 07-APR-2005.
XX
PF 17-SEP-2004; 2004WO-EP010517.
XX
PR 26-SEP-2003; 2003US-0506479P.
PR 20-JAN-2004; 2004US-0537729P.
XX
XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
XX
XX Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
XX
XX WPI; 2005-273382/28.
DR N-PSDB; ADZ18997.
XX
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX
XX Example 2; SEQ ID NO 17; 65pp; English.
XX
XX The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hep106 construct protein
CC used in the scope of the invention.
XX
SQ Sequence 492 AA;

    Query Match      100.0%; Score 104; DB 9; Length 492;
    Best Local Similarity 100.0%; Pred. No. 6.2e-09;
    Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIIFDPK 19
   |||||
Db 89 PAYLRFGGTKTDFLIIFDPK 107

RESULT 13

```

```

AEB87562
ID AEB87562 standard; protein; 493 AA.
XX
AC AEB87562;
XX
DT 06-OCT-2005 (first entry)
XX
DE Human heparanase 65delta15 deletion mutant.
XX
KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;
KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;
KW leukemia; lymphoma; dermatological disease; hematological disease;
KW immune disorder; inflammation; antiinflammatory; renal disease;
KW nephrotropic; endocrine disease; genitourinary disease;
KW autoimmune disease; immunosuppressive; drug screening; muten.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2005071070-A2.
XX
PD 04-AUG-2005.
XX
PF 20-JAN-2005; 2005WO-IL000068.
XX
PR 22-JAN-2004; 2004IL-00160025.
PR 28-JUL-2004; 2004US-00901943.
XX
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
XX
XX Vlodavsky I, Ilan N, Levy-Adam F;
XX
XX WPI; 2005-564219/57.
DR N-PSDB; AEB87561.
XX
XX New amino acid sequences derived from the 50 kDa subunit of heparanase,
PT for treating or inhibiting malignant proliferative, inflammatory, kidney
PT disorder or autoimmune disorder.
XX
XX Claim 105; SEQ ID NO 4; 167pp; English.
XX
XX The present sequence is that of a deletion mutant of human heparanase,
CC denoted 6delta15, which is devoid of amino acid residues 158-171 of the
CC native protein. The recombinant protein is deficient of heparanase
CC endoglycosidase catalytic activity. The invention relates to amino acid
CC sequences derived from the N-terminus region of the 50 kDa subunit of
CC heparanase, particularly in the regions between amino acid residues 158-
CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise
CC heparin-binding domains. The invention also provides an antibody directed
CC to these sequences, in particular the 158-171 peptide, and compositions
CC and uses of this antibody as a heparanase inhibitor. A screening method
CC is provided for specific heparanase inhibitors. Claimed pharmaceutical
CC compositions comprising (i) a peptide derived from the N-terminus region
CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a
CC peptide, or (iii) an antibody which specifically recognizes the peptide
CC are used for the inhibition of heparanase catalytic activity associated
CC with an inflammatory disorder, kidney disease, autoimmune disease,
CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or
CC with a malignant proliferative disorder, especially a solid or non-solid
CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.
XX
SQ Sequence 493 AA;

    Query Match      100.0%; Score 104; DB 9; Length 493;
    Best Local Similarity 100.0%; Pred. No. 6.2e-09;
    Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIIFDPK 19
   |||||
Db 54 PAYLRFGGTKTDFLIIFDPK 72

RESULT 14

```

```

ADZ18999
ID ADZ18999 standard; protein; 495 AA.
XX
AC ADZ18999;
XX
DT 16-JUN-2005 (first entry)
XX
DE Hep109 construct protein.
XX
KW Enzyme engineering; heparanase; hep109; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX
OS Synthetic.
XX
PN WO2005030962-A1.
XX
PD 07-APR-2005.
XX
PF 17-SEP-2004; 2004WO-EP010517.
XX
PR 26-SEP-2003; 2003US-0506479P.
XX
PR 20-JAN-2004; 2004US-0537729P.
XX
PA (RICE-) 1ST RICERCHE BIOL MOLECOLARE ANGELETTI.
XX
PI Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
XX
DR WPI; 2005-273382/28.
DR N-PSDB; ADZ18998.
XX
PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX
PS Example 2; SEQ ID NO 20; 65pp; English.
XX
CC The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hep109 construct protein
CC used in the scope of the invention.
XX
SQ Sequence 495 AA;

Query Match 100.0%; Score 104; DB 9; Length 495;
Best Local Similarity 100.0%; Pred. No. 6.3e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIIDPK 19
DB 89 PAYLRFGGTKTDFLIIDPK 107
|||||
|||||

RESULT 15
AEB87587
ID AEB87587 standard; protein; 497 AA.
XX
AC AEB87587;
XX
DT 06-OCT-2005 (first entry)
XX
DE Human heparanase 65delta10 deletion mutant.
XX
KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;
KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;
KW leukemia; lymphoma; dermatological disease; hematological disease;
KW immune disorder; inflammation; antiinflammatory; renal disease;
KW nephrotropic; endocrine disease; genitourinary disease;

ADZ18999
KW autoimmune disease; immunosuppressive; drug screening; mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2005071070-A2.
XX
PD 04-AUG-2005.
XX
PF 20-JAN-2005; 2005WO-IL000068.
XX
PR 22-JAN-2004; 2004IL-00160025.
XX
PR 28-JUL-2004; 2004US-00901943.
XX
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
XX
PI Vlodavsky I, Ilan N, Levy-Adam F;
XX
DR WPI; 2005-564219/57.
DR N-PSDB; AEB87586.
XX
PT New amino acid sequences derived from the 50 kDa subunit of heparanase,
PT for treating or inhibiting malignant proliferative, inflammatory, kidney
PT disorder or autoimmune disorder.
XX
PS Claim 106; SEQ ID NO 29; 167pp; English.
XX
CC The present sequence is that of a deletion mutant of human heparanase,
CC denoted 65delta10, which is devoid of amino acid residues 270-280 of the
CC native protein. The recombinant protein is deficient of heparanase
CC endoglycosidase catalytic activity. The invention relates to amino acid
CC sequences derived from the N-terminus region of the 50 kDa subunit of
CC heparanase, particularly in the regions between amino acid residues 158-
CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise
CC heparin-binding domains. The invention also provides an antibody directed
CC to these sequences, in particular the 158-171 peptide, and compositions
CC and uses of this antibody as a heparanase inhibitor. A screening method
CC is provided for specific heparanase inhibitors. Claimed pharmaceutical
CC compositions comprising (i) a peptide derived from the N-terminus region
CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a
CC peptide, or (iii) an antibody which specifically recognizes the peptide
CC are used for the inhibition of heparanase catalytic activity associated
CC with an inflammatory disorder, kidney disease, autoimmune disease,
CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or
CC with a malignant proliferative disorder, especially a solid or non-solid
CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.
XX
SQ Sequence 497 AA;

Query Match 100.0%; Score 104; DB 9; Length 497;
Best Local Similarity 100.0%; Pred. No. 6.3e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIIDPK 19
DB 54 PAYLRFGGTKTDFLIIDPK 72
|||||
|||||

RESULT 16
ADZ19000
ID ADZ19000 standard; protein; 501 AA.
XX
AC ADZ19000;
XX
DT 16-JUN-2005 (first entry)
XX
DE HepGS3 construct protein.
XX
KW Enzyme engineering; heparanase; hepGS3; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX
OS Synthetic.

```



XX WO2005030962-A1.  
 XX  
 XX  
 PD 07-APR-2005.  
 XX  
 XX PF 17-SEP-2004; 2004WO-EP010517.  
 XX  
 PR 26-SEP-2003; 2003US-0506479P.  
 PR 20-JAN-2004; 2004US-0537729P.  
 XX  
 XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.  
 XX  
 PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;  
 XX  
 XX WPI; 2005-273382/28.  
 DR N-PSDB; ADZ19001.  
 XX  
 XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.  
 XX  
 PS Example 2; SEQ ID NO 21; 65pp; English.  
 XX  
 CC The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoprotease. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in  
 CC inhibitor screening assays for the development of therapeutics or  
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a hepgs3 construct protein  
 CC used in the scope of the invention.  
 XX  
 XX Sequence 501 AA;  
 SQ  
 Query Match 100.0%; Score 104; DB 9; Length 501;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 PAYLRFGGTKTDFLIIDPK 19  
 Db |||||  
 89 PAYLRFGGTKTDFLIIDPK 107  
 XX  
 RESULT 17  
 ADZ19005  
 ID ADZ19005 standard; protein; 507 AA.  
 XX  
 AC ADZ19005;  
 XX  
 XX 16-JUN-2005 (first entry)  
 DT  
 XX HepGS6 construct protein.  
 DE  
 XX Enzyme engineering; heparanase; hepgs6; metastasis; autoimmune disease;  
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
 KW immunosuppressive; enzyme.  
 XX  
 OS Synthetic.  
 XX  
 XX WO2005030962-A1.  
 XX  
 XX 07-APR-2005.  
 PD  
 XX PF 17-SEP-2004; 2004WO-EP010517.  
 XX  
 PR 26-SEP-2003; 2003US-0506479P.  
 PR 20-JAN-2004; 2004US-0537729P.  
 XX  
 XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.  
 PA Lahm A, Nardella C, Pallaro M, Steinkuhler C;  
 XX  
 XX

DR WPI; 2005-273382/28.  
 DR N-ESDB; ADZ19003.  
 XX  
 XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.  
 XX  
 PS Example 2; SEQ ID NO 26; 65pp; English.  
 XX  
 CC The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoprotease. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in  
 CC inhibitor screening assays for the development of therapeutics or  
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a hepgs6 construct protein  
 CC used in the scope of the invention.  
 XX  
 XX Sequence 507 AA;  
 SQ  
 Query Match 100.0%; Score 104; DB 9; Length 507;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 PAYLRFGGTKTDFLIIDPK 19  
 Db |||||  
 89 PAYLRFGGTKTDFLIIDPK 107  
 XX  
 RESULT 18  
 ADY27058  
 ID ADY27058 standard; protein; 508 AA.  
 XX  
 AC ADY27058;  
 XX  
 XX 05-MAY-2005 (first entry)  
 DT  
 XX Human inactive heparanase protein.  
 DE  
 XX Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
 KW neurological disease; viral infection; infection; cytostatic;  
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
 KW protease; enzyme; enzyme purification.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO2005016227-A2.  
 FN  
 XX 24-FEB-2005.  
 PD  
 XX 12-AUG-2004; 2004WO-IL000744.  
 PF  
 XX 14-AUG-2003; 2003US-0494800P.  
 PR  
 PR 12-JAN-2004; 2004US-0535492P.  
 XX  
 PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.  
 XX  
 XX Van Gelder JM, Miron D;  
 PI  
 XX WPI; 2005-182203/19.  
 DR  
 XX Regulating heparanase activity, useful for treating heparanase-associated  
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases, heparanase  
 PT neurological diseases or viral diseases) comprises modulating heparanase  
 PT activation.  
 XX  
 PS Claim 257; SEQ ID NO 34; 211pp; English.  
 XX  
 XX The invention relates to a method of regulating heparanase activity in a  
 CC tissue or regulating a biological process depending at least in part on  
 CC heparanase activity comprising modulating heparanase activation. The  
 CC invention also relates to methods of treating a heparanase- or heparin

CC binding protein-associated disease or disorder in a subject, a  
 CC pharmaceutical composition for use in the treatment of a heparanase-  
 CC associated disease or disorder comprising a therapeutic amount of an  
 CC agent capable of modulating heparanase activation and a pharmaceutical  
 CC carrier or diluent, a method of identifying a protease activator of  
 CC heparanase, a protease substrate mimetic comprising a peptide  
 CC representing a subset or all substrate residues or cleavage sites of  
 CC human heparanase or an equivalent non-human heparanase, a method of  
 CC producing active heparanase and a method of modulating an adhesion  
 CC activity of heparanase. The composition and methods are useful for  
 CC modulating heparanase activation and for treating heparanase-associated  
 CC diseases or disorders such as cancer, inflammation, cardiovascular  
 CC diseases, neurological diseases or viral infections. This sequence  
 CC represents a human inactive heparanase protein used in the scope of the  
 CC invention.

XX  
 SQ Sequence 508 AA;

Query Match 100.0%; Score 104; DB 9; Length 508;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19  
 |||||  
 Db 54 PAYLRFGGTKTDFLIFDPK 72

# RESULT 19

ADZ19006  
 ID ADZ19006 standard; protein; 526 AA.

XX AC ADZ19006;

XX DT 16-JUN-2005 (first entry)

XX DE HepHyaluro construct protein.

XX Enzyme engineering; heparanase; hephyaluro; metastasis;  
 KW autoimmune disease; inflammation; neoplasm; immune disorder;  
 KW antiinflammatory; cytostatic; immunosuppressive; enzyme.

XX OS Synthetic.

XX PN WO2005030962-A1.

XX PD 07-APR-2005.

XX PF 17-SEP-2004; 2004WO-EP010517.

XX PR 26-SEP-2003; 2003US-0506479P.

XX PR 20-JAN-2004; 2004US-0537729P.

XX PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.

XX PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;

XX DR WPI; 2005-273382/28.

XX DR N-PSDB; ADZ19007.

XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.

XX Example 2; SEQ ID NO 27; 65pp; English.

XX The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoproteinase. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in  
 CC inhibitor screening assays for the development of therapeutics or  
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a hephyaluro construct

CC protein used in the scope of the invention.

SQ Sequence 526 AA;

Query Match 100.0%; Score 104; DB 9; Length 526;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19  
 |||||  
 Db 89 PAYLRFGGTKTDFLIFDPK 107

# RESULT 20

ABB07815

ID ABB07815 standard; protein; 527 AA.

XX AC ABB07815;

XX DT 03-JUL-2002 (first entry)

XX DE Chicken signal peptide/human heparanase chimeric protein sequence.

XX Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;  
 KW anti-protozoan; neuroprotective; heparin; chicken; human; chimeric.

XX OS Synthetic.

XX OS Gallus gallus.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..19

FT Protein /note= "chicken heparanase signal peptide"

FT Protein 20..527

FT Protein /note= "human heparanase mature protein"

XX PN US2002034810-A1.

XX PD 21-MAR-2002.

XX PF 16-AUG-2001; 2001US-00930218.

XX PR 20-SEP-2000; 2000US-00666390.

XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX PI Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;

XX DR WPI; 2002-338926/37.

XX DR N-PSDB; ABL40753.

XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful  
 PT to treat various heparin-related disorders and the signal peptide is  
 PT useful in production of membrane-targeted or secreted recombinant  
 PT proteins.

XX Disclosure; Page 26-28; 39pp; English.

XX The invention relates to an isolated avian and reptile nucleic acid,  
 CC encoding a polypeptide with heparanase catalytic activity. The signal  
 CC peptide of the nucleic acid can be used to express membrane-associated or  
 CC secreted proteins in heterologous expression systems. The encoded  
 CC polypeptides can be used to prevent tumour angiogenesis, metastasis and  
 CC invasion, and to intervene with pathologies associated with impaired  
 CC heparin-binding growth factors, cellular responses to heparin-binding  
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,  
 CC cellular susceptibility to viral, protozoa and bacterial infections or  
 CC disintegration of neurodegenerative plaques. The present sequence  
 CC represents a chicken signal peptide/human heparanase chimeric protein  
 CC sequence

XX Sequence 527 AA;

Query Match 100.0%; Score 104; DB 5; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIIDPK 19  
 |||||  
 Db 73 PAYLRFGGTKTDFLIIDPK 91

RESULT 21  
 ABW02018  
 ID ABW02018 standard; protein; 527 AA.  
 XX  
 AC ABW02018;  
 XX  
 DT 12-FEB-2004 (first entry)  
 XX  
 DE Chimeric human-chicken heparanase protein.  
 XX  
 KW Chicken; heparanase; tumour cell metastasis; inflammation; autoimmunity;  
 KW wound healing; angiogenesis; restenosis; Genstmann-Straussler Syndrome;  
 KW neurodegenerative disease; atherosclerosis; Creutzfeldt-Jakob disease;  
 KW infection; Scrapie; Alzheimer's disease; protein therapy; cytostatic;  
 KW immunosuppressive; vulnery; bactericide; anti-angiogenic; virucide;  
 KW antisclerotic; neuroprotective; protozoacide; chimeric; fusion protein;  
 KW enzyme; human.  
 XX  
 OS Chimeric - Gallus gallus.  
 OS Chimeric - Homo sapiens.  
 XX  
 XX US2003180788-A1.  
 XX  
 PN 25-SEP-2003.  
 XX  
 PD 08-MAY-2003; 2003US-00431438.  
 XX  
 PF 20-SEP-2000; 2000US-00666390.  
 XX  
 PR 16-AUG-2001; 2001US-00930218.  
 XX  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX  
 XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;  
 XX  
 XX WPI; 2003-843931/78.  
 DR N-PSDB; AAD63532.  
 XX  
 PT Recombinant jungle red fowl (Gallus gallus) heparanase protein, useful  
 for treating cancers, microbial infections and aiding wound healing.  
 XX  
 XX Example; Page 26-28; Opp; English.  
 XX  
 CC The present invention relates to novel jungle red fowl heparanase protein  
 CC and polynucleotides encoding such proteins. Heparanase sequences can be  
 CC used to develop treatments for various diseases, to develop diagnostic  
 CC assays for these diseases and to provide new tools for basic and directed  
 CC research especially in the fields of medicine and biology. They can be  
 CC used to develop new drugs to inhibit tumour cell metastasis, inflammation  
 CC and autoimmunity. Recombinant heparanase offers a potential treatment for  
 CC wound healing, angiogenesis, restenosis, atherosclerosis, inflammation,  
 CC neurodegenerative diseases (e.g. Genstmann-Straussler Syndrome, Scrapie,  
 CC Creutzfeldt-Jakob disease and Alzheimer's disease) and certain viral and  
 CC some bacterial and protozoa infections. Recombinant heparanase can also  
 CC be used to neutralise plasma heparin, as a potential replacement of  
 CC protamine. Sequences of the invention are used in protein therapy. The  
 CC present sequence is chimeric human-chicken heparanase protein  
 XX  
 SQ Sequence 527 AA;

Query Match 100.0%; Score 104; DB 7; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIIDPK 19  
 |||||  
 Db 73 PAYLRFGGTKTDFLIIDPK 91

RESULT 22  
 ADO63825  
 ID ADO63825 standard; protein; 527 AA.  
 XX  
 AC ADO63825;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Chimeric heparanase mutant E225A, SEQ ID:10.  
 XX  
 KW Human; chicken; heparanase; heparanase-derived protein;  
 KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;  
 KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;  
 KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; IAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnery; mutant; mutein.  
 XX  
 OS Homo sapiens.  
 OS Gallus gallus.  
 OS Synthetic.  
 OS Chimeric.  
 XX  
 XX Key Location/Qualifiers  
 FH Peptide 1..18  
 FT /note= "Chicken heparanase signal peptide"  
 FT Region 19..527  
 FT /note= "Corresponds to residues 35-543 of human  
 FT heparanase mutant E225A (SEQ ID NO:7)"  
 FT Misc-difference 209  
 FT /note= "Ala replaces wild-type Glu (active site proton  
 FT donor). Corresponds to residue 225 of human heparanase  
 FT mutant E225A (SEQ ID NO:7)"  
 FT Active-site 327  
 FT /note= "Active site nucleophile"  
 XX  
 PN WO2004048558-A2.  
 XX  
 XX 10-JUN-2004.  
 PD  
 XX 24-NOV-2003; 2003WO-IL000989.  
 PF  
 XX 24-NOV-2002; 2002IL-00153059.  
 PR  
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA  
 XX Vlodavsky I, Zcharia E, Goldshmidt O, Ilan N;  
 PI  
 XX WPI; 2004-450373/42.  
 DR N-PSDB; ADO63819.  
 XX  
 PT New nucleic acid construct comprising heparanase-derived polypeptide,  
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 PT thrombasthenia, or Bernard-Soulier syndrome.  
 XX  
 PS Claim 10; SEQ ID NO 10; 128pp; English.  
 XX  
 CC The invention relates to nucleic acid constructs comprising a nucleic  
 CC acid encoding a heparanase-derived protein which lacks heparanase  
 CC endoglycosidase catalytic activity but which retains its cell-cell and  
 CC cell-matrix adhesion properties. The constructs of the invention  
 CC optionally further comprise operably linked regulatory elements. The  
 CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the  
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-

CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
 CC host cells are useful in preparing a tissue sealant composition for  
 CC sealing injuries, reducing the loss of blood, accelerating the healing  
 CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialization of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be used in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte  
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be used to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents a chimeric  
 CC protein comprising the signal peptide of chicken heparanase and residues  
 CC 35-543 of the human heparanase mutant E225A.  
 XX SQ Sequence 527 AA;

Query Match 100.0%; Score 104; DB 8; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19  
 |||||  
 Db 73 PAYLRFGGTKTDFLIFDPK 91

RESULT 23  
 ADO63827  
 ID ADO63827 standard; protein; 527 AA.  
 XX AC ADO63827;  
 XX DT 26-AUG-2004 (first entry)  
 XX DE Chimeric heparanase mutant E225A/E343A, SEQ ID:12.  
 XX KW Human; chicken; heparanase; heparanase-derived protein;  
 KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;  
 KW matrix adhesion; tissue sealant; injury; blood loss; endothelialization;  
 KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; LAD; leukocyte adhesion deficiency; drug screening;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnery; mutant; mutein.  
 XX OS Homo sapiens.  
 OS Gallus gallus.  
 OS Synthetic.  
 OS Chimeric.

Key Location/Qualifiers  
 FT Peptide 1..18 "Chicken heparanase signal peptide"  
 FT Region 19..527  
 FT /note= "Corresponds to residues 35-543 of human  
 FT heparanase mutant E225A/E343A (SEQ ID NO:9)"  
 FT Misc-difference 209  
 FT /note= "Ala replaces wild-type Glu (active site proton  
 FT donor). Corresponds to residue 225 of human heparanase  
 FT mutant E225A/E343A (SEQ ID NO:9)"  
 FT Misc-difference 327  
 FT /note= "Ala replaces wild-type Glu (active site  
 FT nucleophile). Corresponds to residue 343 of human  
 FT heparanase mutant E225A/E343A (SEQ ID NO:9)"  
 XX WO2004048558-A2.  
 XX 10-JUN-2004.  
 XX 24-NOV-2003; 2003WO-IL000989.  
 XX 24-NOV-2002; 2002IL-00153059.

XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
 XX WPI; 2004-450373/42.  
 DR N-PSDB; ADO63821.  
 XX New nucleic acid construct comprising heparanase-derived polypeptide,  
 PT useful in treating wounds, leukocyte Adhesion Deficiency, Glanzmann's  
 PT thrombasthenia, or Bernard-Soulier syndrome.  
 PS Claim 10; SEQ ID NO 12; 128pp; English.  
 XX The invention relates to nucleic acid constructs comprising a nucleic  
 CC acid encoding a heparanase-derived protein which lacks heparanase  
 CC endoglycosidase catalytic activity but which retains its cell-cell and  
 CC cell-matrix adhesion properties. The constructs of the invention  
 CC optionally further comprise operably linked regulatory elements. The  
 CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the  
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
 CC host cells are useful in preparing a tissue sealant composition for  
 CC sealing injuries, reducing the loss of blood, accelerating the healing  
 CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialization of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be used in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte  
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be used to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents a chimeric  
 CC protein comprising the signal peptide of chicken heparanase and residues  
 CC 35-543 of the human heparanase double mutant E225A/E343A.  
 XX SQ Sequence 527 AA;

Query Match 100.0%; Score 104; DB 8; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19  
 |||||  
 Db 73 PAYLRFGGTKTDFLIFDPK 91

RESULT 24  
 ADO63826  
 ID ADO63826 standard; protein; 527 AA.  
 XX AC ADO63826;  
 XX DT 26-AUG-2004 (first entry)  
 XX DE Chimeric heparanase mutant E343A, SEQ ID:11.  
 XX KW Human; chicken; heparanase; heparanase-derived protein;  
 KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;  
 KW matrix adhesion; tissue sealant; injury; blood loss; endothelialization;  
 KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnery; mutant; mutein.  
 XX OS Homo sapiens.  
 OS Gallus gallus.

OS Synthetic.  
 XX Chimeric.  
 XX  
 XX Key Location/Qualifiers  
 FT 1. 18  
 FT Peptide /note= "Chicken heparanase signal peptide"  
 FT 19. 527  
 FT Region /note= "Corresponds to residues 35-543 of human  
 FT heparanase mutant E343A (SEQ ID NO:8)"  
 FT 209  
 FT Active-site /note= "Active site proton donor"  
 FT 327  
 FT Misc-difference /note= "Ala replaces wild-type Glu (active site  
 FT nucleophile). Corresponds to residue 343 of human  
 FT heparanase mutant E343A (SEQ ID NO:8)"  
 XX  
 XX WO2004048558-A2.  
 XX  
 XX 10-JUN-2004.  
 XX  
 XX 24-NOV-2003; 2003WO-IL000989.  
 XX  
 XX 24-NOV-2002; 2002IL-00153059.  
 XX  
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX  
 XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
 XX WPI; 2004-450373/42.  
 DR N-PSDB; ADO63820.  
 XX  
 XX New nucleic acid construct comprising heparanase-derived polypeptide,  
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 PT thrombasthenia, or Bernard-Soulier syndrome.  
 XX  
 XX Claim 10; SEQ ID NO 11; 128pp; English.  
 XX  
 XX The invention relates to nucleic acid constructs comprising a nucleic  
 CC acid encoding a heparanase-derived protein which lacks heparanase  
 CC endoglycosidase catalytic activity but which retains its cell-cell and  
 CC cell-matrix adhesion properties. The constructs of the invention  
 CC optionally further comprise operably linked regulatory elements. The  
 CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the  
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
 CC host cells are useful in preparing a tissue sealant composition for  
 CC sealing injuries, reducing the loss of blood, accelerating the healing  
 CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialisation of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be use in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte  
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents a chimeric  
 CC protein comprising the signal peptide of chicken heparanase and residues  
 CC 35-543 of the human heparanase mutant E343A.  
 XX  
 XX Sequence 527 AA;  
 SQ  
 Query Match 100.0%; Score 104; DB 8; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PAYLRFGGTKTDFLIIDPK 19  
 ||||||||||||||||

Db 73 PAYLRFGGTKTDFLIIDPK 91  
 RESULT 25  
 ADZ19004  
 ID ADZ19004 standard; protein; 527 AA.  
 XX  
 XX ADZ19004;  
 AC  
 XX 16-JUN-2005 (first entry)  
 DT  
 XX HepGS4 construct protein.  
 DE  
 XX Enzyme engineering; heparanase; hepGS4; metastasis; autoimmune disease;  
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
 KW immunosuppressive; enzyme.  
 XX  
 XX Synthetic.  
 OS  
 XX WO2005030962-A1.  
 FN  
 XX 07-APR-2005.  
 PD  
 XX 17-SEP-2004; 2004WO-EP010517.  
 PF  
 XX 26-SEP-2003; 2003US-0506479P.  
 PR  
 XX 20-JAN-2004; 2004US-0537729P.  
 PR  
 XX (RICE-) IST RICERCH BIOL MOLECOLARE ANGELETTI.  
 PA  
 XX Lahm A, Nardella C, Pallaoro M, Steinkuhler C;  
 PI WPI; 2005-273382/28.  
 XX  
 XX N-PSDB; ADZ19002.  
 DR  
 XX  
 XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.  
 XX  
 XX Example 2; SEQ ID NO 25; 65pp; English.  
 PS  
 XX The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoprotease. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in  
 CC inhibitor screening assays for the development of therapeutics or  
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a hepGS4 construct protein  
 CC used in the scope of the invention.  
 XX  
 XX Sequence 527 AA;  
 SQ  
 Query Match 100.0%; Score 104; DB 9; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PAYLRFGGTKTDFLIIDPK 19  
 ||||||||||||||||  
 Db 89 PAYLRFGGTKTDFLIIDPK 107  
 RESULT 26  
 AAY34173  
 ID AAY34173 standard; protein; 530 AA.  
 XX  
 XX AAY34173;  
 AC  
 XX 15-NOV-1999 (first entry)  
 DT  
 XX Human pre-proheparanase protein sequence.  
 DE  
 XX Human; pre-proheparanase; platelet; wound healing; angiogenesis blocker;  
 KW

KW inflammation; psoriasis; diabetic retinopathy; solid tumour; arthritis;  
 KW heparin degradation; anticoagulant neutralisation; asthma; CNS disease;  
 KW inflammatory disease; vascular restenosis; atherosclerosis; diagnosis;  
 KW tumour growth; fibroproliferative disorder; neurodegenerative disease;  
 therapy.

XX Homo sapiens.  
 XX OS  
 PN WO9943830-A2.  
 XX  
 XX 02-SEP-1999.  
 XX  
 XX 18-FEB-1999; 99WO-US001489.  
 XX  
 XX 24-FEB-1998; 98US-0075706P.  
 PR 26-MAR-1998; 98US-0079401P.  
 XX  
 XX (PHAA ) PHARMACIA & UPJOHN CO.  
 XX  
 XX Heinrikson RL, Fairbanks MB, Mildner AM;  
 PI  
 XX WPI; 1999-540598/45.  
 DR N-PSDB; AAZ11236.  
 XX

PT New isolated platelet heparanase polypeptides, used to develop products  
 PT for, e.g. wound healing and blocking angiogenesis.  
 XX

PS Claim 12; Fig 7; 57pp; English.  
 XX  
 XX This sequence is the human pre-proheparanase of the invention. This  
 CC sequence was isolated from human platelets. The heparanase can be used  
 CC for identifying agents which alter heparanase activity. The heparanase  
 CC can be used for wound healing or for blocking angiogenesis or  
 CC inflammation. It can be used for treating e.g. psoriasis, diabetic  
 CC retinopathy or solid tumours, or for the degradation of heparin and the  
 CC neutralisation of heparin's anticoagulant properties during surgery.  
 CC Inhibitors of heparanase activity can be used in the treatment of  
 CC arthritis, asthma, and other inflammatory diseases, vascular restenosis,  
 CC atherosclerosis, tumour growth and progression, fibroproliferative  
 CC disorders, and central nervous system (CNS) and neurodegenerative  
 CC diseases. The products can also be used for detection and diagnosis. The  
 CC purified heparanase, both recombinantly produced human heparanase and  
 CC heparanase isolated from human platelet activity, allows for the  
 CC convenient selection of compounds having anti-heparanase activity, i.e.  
 CC inhibitors of heparanase activity, by measuring inhibition of heparanase  
 CC activity. Inhibition of heparanase activity can be measured by blocking  
 CC heparanase-mediated release of radioactive fragments from in vivo  
 CC radiolabelled (HSPG)/heparin  
 XX

SQ Sequence 530 AA;  
 Query Match 100.0%; Score 104; DB 2; Length 530;  
 Best Local Similarity 100.0%; Pred. No. 6.8e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19  
 |||||  
 Db 76 PAYLRFGGTKTDFLIFDPK 94

RESULT 27  
 AAY17083  
 ID AAY17083 standard; protein; 532 AA.  
 XX  
 AC AAY17083;  
 XX  
 XX 21-JUL-1999 (first entry)  
 DT  
 XX  
 DE Seq ID No: 15 of WO9921975.  
 XX  
 KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;  
 KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;  
 KW atherosclerosis; atherosclerosis; inflammation; tissue development;

KW human; HSPG.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9921975-A1.  
 XX  
 XX 06-MAY-1999.  
 XX  
 XX 28-OCT-1998; 98WO-AU000898.  
 PF  
 XX 28-OCT-1997; 97AU-00000062.  
 PR 09-DEC-1997; 97AU-00000812.  
 XX  
 XX (AUSU ) UNIV AUSTRALIAN NAT.  
 PA  
 XX  
 XX Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;  
 PI  
 XX WPI; 1999-312956/26.  
 DR N-PSDB; AAX37260.  
 XX  
 XX Polynucleotides encoding mammalian endoglucuronidases, especially  
 PT heparanases, useful to promote wound healing.  
 PT  
 XX Claim 6; Page 76-79; 112pp; English.

XX The invention relates to nucleic acid sequences that encode heparanase  
 CC enzymes having endoglucuronidase activity. Recombinant heparanases are  
 CC capable of removing the HS side chain from heparan sulfate proteoglycan  
 CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to  
 CC inhibit heparanase. This is useful for treatment of a physiological or  
 CC medical condition associated with elevated heparanase activity, such as  
 CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,  
 CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and  
 CC rat heparanases can be used to enhance wound healing, especially  
 CC associated with tissue development and repair. The conditions mentioned  
 CC above can be diagnosed using specific antibodies, and also using primers  
 CC and probes specific for the heparanase polynucleotides. Other uses of the  
 CC heparanases include sequencing sulfated molecules such as HSPG  
 XX

SQ Sequence 532 AA;  
 Query Match 100.0%; Score 104; DB 2; Length 532;  
 Best Local Similarity 100.0%; Pred. No. 6.8e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19  
 |||||  
 Db 89 PAYLRFGGTKTDFLIFDPK 107

RESULT 28  
 AAY02345  
 ID AAY02345 standard; protein; 543 AA.  
 XX  
 AC AAY02345;  
 XX  
 XX 09-JUL-1999 (first entry)  
 DT  
 XX  
 DE A human heparanase protein.  
 XX  
 KW Heparanase; hp; modulator; heparin-binding growth factor;  
 KW cellular response; cytokine; cell interaction; plasma lipoprotein;  
 KW cellular susceptibility; infection; disintegration;  
 KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;  
 KW atherosclerosis; inflammation; neurodegenerative disease; neutralase;  
 KW plasma heparin; micrometastasis; autoimmune lesion; renal failure.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9911798-A1.  
 XX  
 PD 11-MAR-1999.  
 XX

PF 31-AUG-1998; 98WO-US017954.  
 XX  
 XX 02-SEP-1997; 97US-00922170.  
 XX 02-JUL-1998; 98US-00109386.  
 PR  
 PA (INST-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA (FRIE/) FRIEDMAN M M.  
 XX  
 XX Pecker I, Vlodavsky I, Feinstein E;  
 XX  
 XX WPI; 1999-302255/25.  
 DR N-PSDB; AAX35648.  
 XX  
 XX New human polynucleotide useful for treating angiogenesis, restenosis,  
 PT and inflammation.  
 PT  
 XX Claim 6; Fig 1; 63pp; English.  
 PS  
 XX The specification describes a polypeptide having heparanase (hp)  
 CC activity. The recombinant protein is used as a modulator of heparin-  
 CC binding growth factors, cellular responses to heparin-binding growth  
 CC factors and cytokines, cell interaction with plasma lipoproteins,  
 CC cellular susceptibility to viral, protozoal and bacterial infections or  
 CC disintegration of neurodegenerative plaques. Heparanase may be useful for  
 CC conditions such as wound healing, angiogenesis, restenosis,  
 CC atherosclerosis, inflammation, neurodegenerative diseases, and viral  
 CC infections. Mammalian heparanase can be used to neutralize plasma  
 CC heparin, and anti-heparanase antibodies may be applied for  
 CC immunodetection and diagnosis of micrometastases, autoimmune lesions, and  
 CC renal failure in biopsy specimens, plasma samples, and body fluids. The  
 CC present sequence represents human heparanase  
 XX  
 XX Sequence 543 AA;  
 SQ  
 Query Match 100.0%; Score 104; DB 2; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PAYLRFGGTKTDFLIFDPK 19  
 DB 89 PAYLRFGGTKTDFLIFDPK 107  
 RESULT 29  
 AAY17082  
 ID AAY17082 standard; protein; 543 AA.  
 XX  
 AC AAY17082;  
 XX  
 DT 21-JUL-1999 (first entry)  
 XX  
 DE Human heparanase enzyme.  
 XX  
 KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;  
 KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;  
 KW arteriosclerosis; atherosclerosis; inflammation; tissue development;  
 KW human; HSPG.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO9921975-A1.  
 PN  
 XX 06-MAY-1999.  
 PD  
 XX 28-OCT-1998; 98WO-AU0000898.  
 PF  
 XX 28-OCT-1997; 97AU-00000062.  
 PR  
 PR 09-DEC-1997; 97AU-00000812.  
 XX  
 XX (AUSU ) UNIV AUSTRALIAN NAT.  
 PA  
 XX Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;  
 PI

XX WPI; 1999-312956/26.  
 DR N-PSDB; AAX37259.  
 XX  
 XX Polynucleotides encoding mammalian endoglucuronidases, especially  
 PT heparanases, useful to promote wound healing.  
 XX  
 PS Claim 6; Page 69-73; 112pp; English.  
 XX  
 XX The invention relates to nucleic acid sequences that encode heparanase  
 CC enzymes having endoglucuronidase activity. Recombinant heparanases are  
 CC capable of removing the HS side chain from heparan sulfate proteoglycan  
 CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to  
 CC inhibit heparanase, this is useful for treatment of a physiological or  
 CC medical condition associated with elevated heparanase activity, such as  
 CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,  
 CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and  
 CC rat heparanases can be used to enhance wound healing, especially  
 CC associated with tissue development and repair. The conditions mentioned  
 CC above can be diagnosed using specific antibodies, and also using primers  
 CC and probes specific for the heparanase polynucleotides. Other uses of the  
 CC heparanases include sequencing sulfated molecules such as HSPG. The  
 CC present sequence represents a human heparanase  
 XX  
 XX Sequence 543 AA;  
 SQ  
 Query Match 100.0%; Score 104; DB 2; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PAYLRFGGTKTDFLIFDPK 19  
 DB 89 PAYLRFGGTKTDFLIFDPK 107  
 RESULT 30  
 AAY57590  
 ID AAY57590 standard; protein; 543 AA.  
 XX  
 AC AAY57590;  
 XX  
 DT 02-MAR-2000 (first entry)  
 XX  
 DE Human heparanase.  
 XX  
 KW Human; heparanase; hpa; genetic modification; expression; anticancer;  
 KW angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumour;  
 KW anti-atherosclerotic; anti-inflammatory; antineurodegeneration;  
 KW heparan sulphate; heparin-binding growth factor; tumour angiogenesis;  
 KW metastasis; wound healing; restenosis; atherosclerosis; inflammation;  
 KW neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;  
 KW micrometastasis; autoimmune lesion; kidney failure.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO9957244-A1.  
 PN  
 XX 11-NOV-1999.  
 PD  
 XX 29-APR-1999; 99WO-US009256.  
 PF  
 XX 01-MAY-1998; 98US-00071618.  
 PR  
 PR 02-MAR-1999; 99US-00260038.  
 XX  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (FRIE/) FRIEDMAN M M.  
 XX  
 XX Ben-Artzi H, Ayal-Hershkovitz M, Yacoby-Zeevi O, Pecker I;  
 PI Peleg Y, Shlomi Y;  
 XX  
 XX WPI; 2000-062144/05.  
 DR N-PSDB; AAZ39195.  
 XX



Engineered cells that express recombinant heparanase, useful therapeutically, e.g. for treating angiogenesis and to screen for specific inhibitors, potential anticancer agents.

PS Claim 3; Page 107-109; 118pp; English.

The present invention describes genetically modified cells (A) containing a polynucleotide (I) that encodes a polypeptide with heparanase activity, and express recombinant heparanase (II). Heparanase cleaves heparan sulphate (HS) at specific intrachain sites, resulting in release of heparin-binding growth factors, enzymes and proteins that are sequestered by HS in basement membranes, extracellular matrix or cell surfaces. It may also be implicated in tumour angiogenesis and metastases. (II) is potentially useful in wound healing and for treating angiogenesis, restenosis, atherosclerosis, inflammation, neurodegeneration, viral infection and cystic fibrosis. It can also be used to neutralise heparin (an alternative to protamine) and to screen for specific inhibitors (potentially useful for treating cancer and metastases). Antibodies raised against (II) are used for immunodetection and diagnosis of micrometastases, autoimmune lesions and kidney failure. (A) provide in large quantities, in a form that is homogeneously processed and activated/neutralised by a dedicated protease. The present sequence represents human heparanase

Sequence 543 AA;

Query Match 100.0%; Score 104; DB 3; Length 543;  
Best Local Similarity 100.0%; Pred. NO. 6.9e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels

Qy	1	PAYLRFGGTKTDFLI	19
Db	89	PAYLRFGGTKTDFLI	107

RESULT 31

AAB08849  
ID AAB08849 standard; protein; 543 AA.

AC AAB08849;

15-JAN-2001 (first entry)

XX DE Amino acid sequence of a human heparanase polypeptide.

Human; heparanase; gene therapy; tumour; inflammation; autoimmunity; heparin-binding growth factor; cytokine; neurodegenerative plaque; wound healing; infection; burn; angiogenesis; restenosis; atherosclerosis; inflammation; neurodegenerative disease; Gerstmann-Sträussler Syndrome; Creutzfeldt-Jakob disease.

XX  
OS  
Homo sapiens.

XX  
PN WO200052178-A1.XX  
PD  
08-SEP-2000.

14-FEB-2000: 2000WO-US003542.

XX  
PR 01-MAR-1999: 99US-00258892.

PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
PA (FRIE/) FRIEDMAN M M.

XX  
PI  
Pecker I. Vlodavsky I. Feinstein E:

XX  
DR WPI; 2000-579289/54.  
DR N-PSDB; AAA75051.

xx New polynucleotides encoding a polypeptide having heparanase activity, PT  
PT useful in wound healing and in gene therapy, particularly in treating  
PT tumor, inflammation, autoimmunity, neurodegenerative diseases.

Claim 22; Fig 1; 152pp; English.

The present sequence represents a human protein with heparanase catalytic activity. The heparanase (hpa) polynucleotide is useful in gene therapy, particularly in treating tumour, inflammation or autoimmunity. Particularly, the polynucleotide is useful in modulating the bioavailability of heparin-binding growth factors, cellular responses to heparin-binding growth factors (e.g. bFGF) and cytokines (e.g. interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular susceptibility to certain viral and some bacterial and protozoa infections, or disintegration of neurodegenerative plaques. The polynucleotide is also useful in wound healing (e.g. thermal, chemical or radiation burns), and in the treatment of angiogenesis, restenosis, atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-Sträussler Syndrome or Creutzfeldt-Jakob disease), and some viral, bacterial or protozoa infections.

Sequence 543 AA;

Query Match 100.0%; Score 104; DB 3; Length 543;  
Best Local Similarity 100.0%; Pred. No. 6.9e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels

Qy	1	PAYL	FGG	TKT	DFLI	FDPK	19
Db	89	PAYL <td>FGG <td>TKT <td>DFLI <td>FDPK</td> <td>107</td> </td></td></td>	FGG <td>TKT <td>DFLI <td>FDPK</td> <td>107</td> </td></td>	TKT <td>DFLI <td>FDPK</td> <td>107</td> </td>	DFLI <td>FDPK</td> <td>107</td>	FDPK	107

RESULT 32

**AAY52990**  
ID AAY52990 standard; protein; 543 AA.

AC AAY52990;

DT 21-FEB-2000 (first entry)

Human heparanase protein sequence.

Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;  
antiadipatic; immunomodulatory; anti-inflammatory; nephrotropic;  
KW metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;  
KW mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;  
KW inflammation; haemorrhagic nephritis; nephrotic syndrome;  
KW autoimmune disease; anti-cancer; kidney disease.

XX Homo sapiens.

XX  
PN  
W09957153-A1.

11-NOV-1999.  
PD  
XX

XX 29-APR-1999: 99WO-US009255.

01-MAY-1998: 98US-00071739.

AA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
PA (FRIE/) FRIEDMAN M M.

XX  
PI Pecker I, vlodavsky I, Friedman Y, Perets T;

XX  
DR WPI; 2000-052944/04.  
DR N-PSDB; AAZ33290.

Heparanase-specific molecular probes useful for diagnosis and treatment, e.g. of tumors, and for targeted drug delivery.

PS Example; Page 81-82; 90pp; English.

The present invention describes heparanase-specific molecular probes, useful for methods of detecting heparanase in situ. The probes and anti-heparanase antibodies are used to detect or quantify the expression of

CC heparanase, for diagnosis and monitoring of diseases (especially  
 CC metastasis), for treatment of heparanase-associated diseases (e.g.  
 CC tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,  
 CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer (or its  
 CC metastases) derived from liver, prostate, bladder, breast, ovary, cervix,  
 CC colon, skin, intestine, stomach, uterus and pancreas, kidney disease,  
 CC diabetes and inflammation, haemorrhagic nephritis, nephrotic syndrome,  
 CC sepsis and inflammatory or autoimmune disease), for targeted drug  
 CC delivery (e.g. of anticancer agents) and as research reagents. The  
 CC present sequence represents human heparanase, which is used in the  
 CC exemplification of the present invention  
 XX  
 XX Sequence 543 AA;

Query Match 100.0%; Score 104; DB 3; Length 543;

Best Local Similarity 100.0%; Pred. No. 6.9e-09; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 0;

Qy 1 PAYLRFGGTKTDFLIIDPK 19  
 |||||  
 Db 89 PAYLRFGGTKTDFLIIDPK 107

# RESULT 33

AA97635  
 ID AA97635 standard; protein; 543 AA.

XX  
 AC AA97635;

XX 20-APR-2001 (first entry)

XX Human heparanase protein sequence.

XX Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;  
 KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;  
 KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;  
 KW gene therapy; human.

XX Homo sapiens.

OS WO200100643-A2.

PN 04-JAN-2001.

XX 19-JUN-2000; 2000WO-IL000358.

XX 25-JUN-1999; 99US-0140801P.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX Pecker I, Michal I, Itzhaki H;

XX WPI; 2001-137930/14.

XX New polynucleotides and polypeptides that are distantly homologous to  
 PT heparanase, useful in wound healing, as well as in gene therapy protocols  
 PT for angiogenesis, restenosis, atherosclerosis, or inflammation.

XX Disclosure; Page 64-65; 67pp; English.

XX This sequence represents a heparanase of the invention. The heparanase  
 CC DNA and protein sequences are useful in wound healing, angiogenesis,  
 CC restenosis, atherosclerosis, inflammation, pulmonary diseases,  
 CC neurodegenerative diseases (such as Scrape, Alzheimer's disease, and  
 CC Creutzfeldt-Jakob disease) or viral infections. The heparanase coding  
 CC sequence is particularly useful in gene therapy

XX Sequence 543 AA;

Query Match 100.0%; Score 104; DB 4; Length 543;

Best Local Similarity 100.0%; Pred. No. 6.9e-09;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFLIIDPK 19  
 |||||  
 Db 89 PAYLRFGGTKTDFLIIDPK 107

# RESULT 34

AA86206  
 ID AA86206 standard; protein; 543 AA.

XX  
 AC AA86206;

XX 24-AUG-2001 (first entry)

XX Human heparanase inhibitor protein.

XX Heparanase; inhibitor; cardiac insufficiency; cardiant; nephrotropic;  
 KW hepatotropic; veterinary medicine; congestive heart failure; dyspnoea;  
 KW primary cardiomyopathy; peripheral odema; pulmonary congestion;  
 KW hepatic congestion; hydrothorax; ascite; nocturia; human.

OS Homo sapiens.

XX DE19955803-A1.

XX 23-MAY-2001.

XX 19-NOV-1999; 99DE-01055803.

XX 19-NOV-1999; 99DE-01055803.

XX (KNOL) KNOLL AG.

XX Herr D, Hahn A, Laux V;

XX WPI; 2001-368371/39.

XX N-PSDB; AAH20940.

XX Treatment or prevention of cardiac insufficiency and related conditions,  
 PT e.g. pulmonary congestion and dyspnoea, comprises administration of  
 PT heparanase inhibitor.

XX Disclosure; Page 11-13; 16pp; German.

XX This invention describes a novel heparanase inhibitor which can be used  
 CC for the treatment or prevention of cardiac insufficiency and associated  
 CC indications, symptoms and/or malfunctions. The heparanase inhibitor of  
 CC the invention has cardiant, nephrotropic and hepatotropic activity. The  
 CC products of the invention can be used in human and veterinary medicine,  
 CC for the treatment or prevention of congestive heart failure e.g. primary  
 CC cardiomyopathy. Associated conditions treated or prevented with the  
 CC inhibitor are especially peripheral odemas, pulmonary and hepatic  
 CC congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g.  
 CC nocturia can also be treated. This sequence represents the human  
 CC heparanase protein described in the method of the invention

XX Sequence 543 AA;

Query Match 100.0%; Score 104; DB 4; Length 543;

Best Local Similarity 100.0%; Pred. No. 6.9e-09;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFLIIDPK 19  
 |||||  
 Db 89 PAYLRFGGTKTDFLIIDPK 107

# RESULT 35

AA88361  
 ID AA88361 standard; protein; 543 AA.

XX  
 AC AA88361;

XX 23-MAY-2001 (first entry)

XX Human membrane or secretory protein clone PSEC0090.  
XX  
XX Human; secretory protein; membrane protein; vaccine; gene therapy;  
KW rheumatoid arthritis; diabetes.  
XX  
XX Homo sapiens.  
OS  
PN EPI067182-A2.  
XX  
PD 10-JAN-2001.  
XX  
PF 07-JUL-2000; 2000EP-00114090.  
XX  
PR 08-JUL-1999; 99JP-00194179.  
PR 11-JAN-2000; 2000JP-00118775.  
PR 02-MAY-2000; 2000JP-00183766.  
XX  
XX (HELI-) HELIX RES INST.  
PA  
XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
PI WPI; 2001-093989/11.  
XX N-PSDB; AAF93788.  
DR  
XX Nucleic acids encoding secretory proteins/membrane proteins, useful in  
PT gene therapy or as candidate target molecules in drug development.  
PT  
XX Claim 1; SEQ ID NO 90; 609pp + Sequence Listing; English.  
PS  
XX This invention relates to nucleic acid sequences AAF93744 - AAF93916  
CC which encode human secretory or membrane proteins represented by AAF88317  
CC - AAF88419. Included in the invention are primers AAF93917 - AAF94295 and  
CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the  
CC invention. The invention also includes methods for the production of  
CC antibodies directed against the proteins, and cDNA sequences, which can  
CC be used in vaccines. The polynucleotide sequences can be used in gene  
CC therapy. The polynucleotide sequences and the proteins they encode may be  
CC used in the prevention, treatment and diagnosis of diseases associated  
CC with inappropriate secretory protein/membrane protein expression. The  
CC nucleic acids and complementary sequences may also be used as DNA probes  
CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect  
CC and quantitate the presence of similar nucleic acid sequences in samples.  
CC They may also be used to study the expression and function of secretory  
CC proteins/membrane polypeptides and their role in metabolism. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC against them and in assays to identify modulators (agonists and  
CC antagonists) of expression and activity. The antibodies and antagonists  
CC may also be used as therapeutic agents to down regulate expression and  
CC activity. The antibodies may also be used as diagnostic agents for  
CC detecting the presence of the polypeptides in samples (e.g. by enzyme  
CC linked immunosorbant assay (ELISA)). Examples of diseases which may be  
CC treated include rheumatoid arthritis and diabetes  
XX  
SQ Sequence 543 AA;  
  
Query Match 100.0%; Score 104; DB 4; Length 543;  
Best Local Similarity 100.0%; Pred. No. 6.9e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 PAYLRFGGTKTDFLIFDPK 19  
Db 89 PAYLRFGGTKTDFLIFDPK 107  
|||||  
  
RESULT 36  
ABB07813  
ID ABB07813 standard; protein; 543 AA.  
XX  
AC ABB07813;  
XX  
XX 03-JUL-2002 (first entry)  
DT  
XX

DE Human heparanase sequence.  
XX  
KW Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;  
KW anti-protozoan; neuroprotective; heparin; human.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..35  
FT /note= "signal peptide"  
FT Protein 36..543  
FT /note= "mature protein"  
XX  
XX US2002034810-A1.  
PN  
XX 21-MAR-2002.  
PD  
XX 16-AUG-2001; 2001US-00930218.  
PF  
XX 20-SEP-2000; 2000US-00666390.  
PR  
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
PA  
XX Goldshmidt O, Pecker I, Vlodaysky I, Michal I, Zcharia E;  
PI WPI; 2002-338926/37.  
XX  
XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful  
PT to treat various heparin-related disorders and the signal peptide is  
PT useful in production of membrane-targeted or secreted recombinant  
PT proteins.  
XX  
XX Disclosure; Fig 1a; 39pp; English.  
XX  
XX The invention relates to an isolated avian and reptile nucleic acid,  
CC encoding a polypeptide with heparanase catalytic activity. The signal  
CC peptide of the nucleic acid can be used to express membrane-associated or  
CC secreted proteins in heterologous expression systems. The encoded  
CC polypeptides can be used to prevent tumour angiogenesis, metastasis and  
CC invasion, and to intervene with pathologies associated with impaired  
CC heparin-binding growth factors, cellular responses to heparin-binding  
CC growth factors and cytokines, cell interaction with plasma lipoproteins,  
CC cellular susceptibility to viral, protozoa and bacterial infections or  
CC disintegration of neurodegenerative plaques. The present sequence  
CC represents a human heparanase protein sequence used in similarity studies  
XX  
SQ Sequence 543 AA;  
  
Query Match 100.0%; Score 104; DB 5; Length 543;  
Best Local Similarity 100.0%; Pred. No. 6.9e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 PAYLRFGGTKTDFLIFDPK 19  
Db 89 PAYLRFGGTKTDFLIFDPK 107  
|||||  
  
RESULT 37  
ADD18950  
ID ADD18950 standard; protein; 543 AA.  
XX  
AC ADD18950;  
XX  
XX 15-JAN-2004 (first entry)  
DT  
XX Human disease related protein SeqID439.  
DE  
XX human; disease state; cytostatic; antiinflammatory; ophthalmological;  
KW antiarteriosclerotic; vulnery; gene therapy;  
KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;  
KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;  
KW glucose transportation; catecholamine synthesis; iron transport;  
KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;  
XX



DR WPI; 2004-070610/07.  
XX  
PT New antisense oligonucleotide hybridizable with a polynucleotide encoding  
PT a polypeptide with heparanase activity, useful for treating diseases such  
PT as cancer and autoimmune disorders.  
XX  
PS Claim 3; SEQ ID NO 10; 108pp; English.  
XX  
XX The invention relates to an antisense oligonucleotide (ASO) comprising a  
CC polynucleotide or a polynucleotide analogue of at least 10 bases being  
CC hybridisable in vivo, under physiological conditions, with a portion of  
CC a polynucleotide strand encoding a polypeptide having heparanase  
CC catalytic activity. Also included are a method of in vivo downregulating  
CC heparanase activity (comprising administering the ASO in vivo), a method  
CC of treating a subject suffering from a pathological condition  
CC (characterised by heparanase activity, comprising administering ASO to  
CC the subject), a pharmaceutical composition comprising the ASO and a  
CC carrier, an antisense nucleic acid construct (comprising a promoter  
CC sequence and a polynucleotide sequence directing the synthesis of an  
CC antisense RNA sequence of at least 10 bases being hybridisable in vivo,  
CC under physiological conditions, with a polynucleotide strand encoding a  
CC polypeptide having heparanase catalytic activity), a method of in vivo  
CC downregulating heparanase activity (comprising administering in vivo the  
CC antisense nucleic acid construct), a pharmaceutical composition  
CC comprising the antisense nucleic acid construct and a carrier, and an  
CC antisense oligonucleotide comprising a polynucleotide or a polynucleotide  
CC analogue of at least 10 bases being hybridisable in vivo, under  
CC physiological conditions, with a portion of a polynucleotide strand being  
CC characterised by forming at least a portion of an untranslated region  
CC (UTR) for a polynucleotide strand encoding a polypeptide having  
CC heparanase catalytic activity. The methods and compositions of the  
CC present invention are useful for the prevention and/or treatment of  
CC diseases or conditions associated with aberrant heparanase activity, such  
CC as heparanase-dependent cancer, cancer, autoimmune reaction and  
CC inflammation. The gene for human heparanase is located on chromosome 4.  
XX The present sequence is a human heparanase protein.  
XX  
SQ Sequence 543 AA;  
Query Match 100.0%; Score 104; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 6.9e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PAYLRFGGTKTDFLIFDPK 19  
Db |||||  
89 PAYLRFGGTKTDFLIFDPK 107  
RESULT 40  
ADK52086  
ID ADK52086 standard; protein; 543 AA.  
XX  
AC ADK52086;  
XX  
XX 20-MAY-2004 (first entry)  
XX Human atopic dermatitis/psoriasis-associated protein #1.  
XX  
KW Human; atopic dermatitis; psoriasis; dermatological; anti-inflammatory;  
KW antipsoriatic; rash.  
XX  
OS Homo sapiens.  
XX  
XX WO2004016785-A1.  
PN  
XX  
XX 26-FEB-2004.  
XX  
XX 06-AUG-2003; 2003WO-JP009999.  
PF  
XX  
XX 06-AUG-2002; 2002JP-00229319.  
PR  
XX 14-MAY-2003; 2003JP-00136544.  
PR  
XX (GENO-) GENOX RES INC.  
PA

PA (UYJU-) UNIV JUNTENDO.  
XX  
PI Itoh M, Ogawa K, Shinagawa A, Sudo H, Ogawa H, Ra C;  
PI Mitsuishi K;  
XX  
XX WPI; 2004-214514/20.  
DR N-PSDB; ADK51968.  
XX  
XX Detecting atopic dermatitis or psoriasis comprises assaying levels of  
PT expression of an indicator gene at a rash site and non-rash site of a  
PT person with atopic dermatitis or psoriasis.  
XX  
PS Example 2; SEQ ID NO 119; 484pp; Japanese.  
XX  
XX The invention relates to detecting atopic dermatitis or psoriasis  
CC comprising assaying the levels of expression of an indicator gene at a  
CC rash site and non-rash site of a person with atopic dermatitis or  
CC psoriasis, comparing these levels with those of a healthy person, and  
CC determining that if the levels of indicators are higher or lower, then  
CC this indicates the disease. Also included are a reagent for detecting  
CC atopic dermatitis or psoriasis, a kit for screening for treatments, a  
CC transgenic non human vertebrate animal models for the diseases, an agent  
CC for inducing the diseases in mice and a DNA chip for assaying for the  
CC indicator genes. The method is used for treatment, detection and animal  
CC models for research of atopic dermatitis and psoriasis. The present  
CC sequence is a protein encoded by an indicator gene of the invention.  
XX  
SQ Sequence 543 AA;  
Query Match 100.0%; Score 104; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 6.9e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PAYLRFGGTKTDFLIFDPK 19  
Db |||||  
89 PAYLRFGGTKTDFLIFDPK 107  
RESULT 41  
ADM48716  
ID ADM48716 standard; protein; 543 AA.  
XX  
AC ADM48716;  
XX  
XX 03-JUN-2004 (first entry)  
XX Human hpa protein #1.  
XX  
XX Transgenic animal; heparanase; cancer; viral infection; restenosis;  
KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;  
KW human.  
XX  
OS Homo sapiens.  
XX  
XX US2003217375-A1.  
PN  
XX  
XX 20-NOV-2003.  
XX  
XX 24-FEB-2003; 2003US-00371218.  
PF  
XX  
XX 31-AUG-1998; 98WO-US017954.  
PR  
XX 01-MAR-1999; 99US-00258892.  
PR  
XX 06-FEB-2001; 2001US-00776874.  
PR  
XX 19-NOV-2001; 2001US-00988113.  
XX  
XX (ZCHA/) ZCHARIA E.  
PA (VLOD/) VLODAVSKY I.  
PA (METZ/) METZGER S.  
PA (PECK/) PECKER I.  
PA (ILAN/) ILAN N.  
PA (CHAJ/) CHAJEK-SHAUL T.  
PA (GOLD/) GOLDSCHMIDT O.  
XX

PI Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;  
 PI Chajek-Shaul T, Goldshmidt O;  
 XX  
 DR WPI; 2004-021918/02.  
 DR N-PSDB; ADM48715, ADM48717.  
 XX  
 PT New transgenic non-human animal expressing heparinase, useful as models  
 PT for human disease, such as cancers, viral infection, neurodegenerative  
 PT diseases, restenosis, atherosclerosis and pulmonary disorders.  
 XX  
 PS Example 1; SEQ ID NO 10; 106pp; English.  
 XX  
 CC The present invention relates to a transgenic non-human animal whose  
 CC genome comprises an exogenous polynucleotide sequence, including a  
 CC promoter active in tissues of the non-human, a region encoding a human  
 CC heparanase, where the promoter and the region encoding human heparanase  
 CC are operably linked in the exogenous polynucleotide such that human  
 CC heparanase is expressed in at least a portion of the cells of the non-  
 CC human animal. The methods and compositions of the present invention are  
 CC useful for the production of transgenic animals expressing heparanase, to  
 CC be used as models for human diseases such as cancers, viral infection,  
 CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary  
 CC disorders. The present sequence is human hpa protein used in the  
 CC exemplification of the invention.  
 XX  
 SQ Sequence 543 AA;  
 Query Match 100.0%; Score 104; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PAYLRFGGTKTDFLIIDPK 19  
 DB 89 PAYLRFGGTKTDFLIIDPK 107  
 RESULT 42  
 ADM48759  
 ID ADM48759 standard; protein; 543 AA.  
 XX  
 AC ADM48759;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE Human hpa protein #2.  
 XX  
 KW Transgenic animal; heparanase; cancer; viral infection; restenosis;  
 KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;  
 KW human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2003217375-A1.  
 XX  
 PD 20-NOV-2003.  
 XX  
 PF 24-FEB-2003; 2003US-00371218.  
 XX  
 PR 31-AUG-1998; 98WO-US017954.  
 PR 01-MAR-1999; 99US-00258892.  
 PR 06-FEB-2001; 2001US-00776874.  
 PR 19-NOV-2001; 2001US-00988113.  
 XX  
 PA {ZCHA/} ZCHARIA E.  
 PA {VLOD/} VLODAVSKY I.  
 PA {METZ/} METZGER S.  
 PA {PECK/} PECKER I.  
 PA {ILAN/} ILAN N.  
 PA {CHAJ/} CHAJEK-SHAUL T.  
 PA {GOLD/} GOLDSHMIDT O.  
 XX  
 PI Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;  
 PI Chajek-Shaul T, Goldshmidt O;

XX WPI; 2004-021918/02.  
 DR N-PSDB; ADM48748.  
 XX  
 PT New transgenic non-human animal expressing heparinase, useful as models  
 PT for human disease, such as cancers, viral infection, neurodegenerative  
 PT diseases, restenosis, atherosclerosis and pulmonary disorders.  
 XX  
 PS Example 10; Fig 16; 106pp; English.  
 XX  
 CC The present invention relates to a transgenic non-human animal whose  
 CC genome comprises an exogenous polynucleotide sequence, including a  
 CC promoter active in tissues of the non-human, a region encoding a human  
 CC heparanase, where the promoter and the region encoding human heparanase  
 CC are operably linked in the exogenous polynucleotide such that human  
 CC heparanase is expressed in at least a portion of the cells of the non-  
 CC human animal. The methods and compositions of the present invention are  
 CC useful for the production of transgenic animals expressing heparanase, to  
 CC be used as models for human diseases such as cancers, viral infection,  
 CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary  
 CC disorders. The present sequence is human hpa protein used in the  
 CC exemplification of the invention.  
 XX  
 SQ Sequence 543 AA;  
 Query Match 100.0%; Score 104; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PAYLRFGGTKTDFLIIDPK 19  
 DB 89 PAYLRFGGTKTDFLIIDPK 107  
 RESULT 43  
 ADM05074  
 ID ADM05074 standard; protein; 543 AA.  
 XX  
 AC ADM05074;  
 XX  
 DT 01-JUL-2004 (first entry)  
 XX  
 DE Antipsoriatic protein sequence #716.  
 XX  
 KW antipsoriatic; gene therapy; psoriasis; diagnosis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004028479-A2.  
 XX  
 PD 08-APR-2004.  
 XX  
 PF 25-SEP-2003; 2003WO-US030907.  
 XX  
 PR 25-SEP-2002; 2002US-0414006P.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;  
 PI Wu TD;  
 XX  
 DR WPI; 2004-305105/28.  
 DR N-PSDB; ADM05073.  
 XX  
 PA New PRO nucleic acid or polypeptide, useful for preparing a  
 PT pharmaceutical composition for diagnosing or treating psoriasis in a  
 PT mammal.  
 XX  
 PS Claim 9; SEQ ID NO 1468; 3069pp; English.  
 XX  
 CC The invention relates to novel polynucleotide and polypeptides for  
 CC treating psoriasis or a sequence having at least 80% identity to the  
 CC above sequences. The nucleic acid is useful for preparing a composition

CC for diagnosing or treating psoriasis in a mammal. This sequence  
 CC corresponds to one of the polypeptides of the invention.

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 104; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-09;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19  
 |||||  
 Db 89 PAYLRFGGTKTDFLIFDPK 107

# RESULT 44

ADN04902  
 ID ADN04902 standard; protein; 543 AA.

XX AC

ADN04902;

XX DT

01-JUL-2004 (first entry)

XX DE

Antipsoriatic protein sequence #631.

XX KW

antipsoriatic; gene therapy; psoriasis; diagnosis.

XX OS

Homo sapiens.

XX PN

WO2004028479-A2.

XX PD

08-APR-2004.

XX PF

25-SEP-2003; 2003WO-US030907.

XX PR

25-SEP-2002; 2002US-0414006P.

XX PA

(GETH ) GENENTECH INC.

XX PI

Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;

PI Wu TD;

XX DR

WPI; 2004-305105/28.

XX DR

N-PSDB; ADN04901.

XX PT

New PRO nucleic acid or polypeptide, useful for preparing a

PT pharmaceutical composition for diagnosing or treating psoriasis in a

PT mammal.

XX PS

Claim 9; SEQ ID NO 1296; 3069pp; English.

XX CC

The invention relates to novel polynucleotide and polypeptides for

CC treating psoriasis or a sequence having at least 80% identity to the

CC above sequences. The nucleic acid is useful for preparing a composition

CC for diagnosing or treating psoriasis in a mammal. This sequence

CC corresponds to one of the polypeptides of the invention.

XX SQ

Sequence 543 AA;

Query Match 100.0%; Score 104; DB 8; Length 543;

Best Local Similarity 100.0%; Pred. No. 6.9e-09;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19

|||||

Db 89 PAYLRFGGTKTDFLIFDPK 107

# RESULT 45

AD063831

ID AD063831 standard; protein; 543 AA.

XX AC

AD063831;

XX

DT 26-AUG-2004 (first entry)  
 DE Human heparanase mutant E378A.  
 XX  
 KW Human; heparanase; heparanase-derived protein; heparanase mutant;  
 KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;  
 KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;  
 KW vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnery; mutant; muten; enzyme.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Active-site 225  
 FT /note= "Active site proton donor"  
 FT Active-site 343  
 FT /note= "Active site nucleophile"  
 FT Misc-difference 378  
 FT /note= "Ala replaces wild-type Glu"  
 XX  
 PN WO2004048558-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 24-NOV-2003; 2003WO-IL000989.  
 XX  
 PR 24-NOV-2002; 2002IL-00153059.  
 XX  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX  
 PI Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
 XX WPI; 2004-450373/42.  
 DR  
 XX New nucleic acid construct comprising heparanase-derived polypeptide,  
 PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 PT thrombasthenia, or Bernard-Soulier syndrome.  
 XX  
 PS Example 4; Page; 128pp; English.  
 CC  
 CC The invention relates to nucleic acid constructs comprising a nucleic  
 CC acid encoding a heparanase-derived protein which lacks heparanase  
 CC endoglycosidase catalytic activity but which retains its cell-cell and  
 CC cell-matrix adhesion properties. The constructs of the invention  
 CC optionally further comprise operably linked regulatory elements. The  
 CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the  
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
 CC are replaced with Ala (ADO63823-ADO63824), and the proteins may  
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
 CC host cells are useful in preparing a tissue sealant composition for  
 CC sealing injuries, reducing the loss of blood, accelerating the healing  
 CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialisation of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be used in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte  
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be used to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents a human  
 CC heparanase mutant E378A created in an example of the invention which  
 CC retains its heparanase catalytic activity. The present sequence is not  
 CC shown in the invention, but is derived from the protein sequence of  
 CC GenBank accession number AF144325 and the information provided on page  
 CC 70.



XX Sequence 543 AA;  
SQ

Query Match 100.0%; Score 104; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 6.9e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFLIIDPK 19  
Db 89 PAYLRFGGTKTDFLIIDPK 107

RESULT 46  
ADO63824  
ID ADO63824 standard; protein; 543 AA.  
XX  
AC ADO63824;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE Human heparanase mutant E225A/E343A, SEQ ID:9.  
XX  
KW Human; heparanase; heparanase-derived protein; heparanase mutant;  
KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;  
KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;  
KW vascular graft; platelet adhesion; platelet aggregation;  
KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
KW vulnerary; mutant; mutein.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 225  
FT /note= "Ala replaces wild-type Glu (active site proton donor)"  
FT  
FT Misc-difference 343  
FT /note= "Ala replaces wild-type Glu (active site nucleophile)"  
XX  
PN WO2004048558-A2.  
XX  
PD 10-JUN-2004.  
XX  
PF 24-NOV-2003; 2003WO-IL000989.  
XX  
PR 24-NOV-2002; 2002IL-00153059.  
XX  
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
XX  
PI Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
XX  
DR WPI: 2004-450373/42.  
DR N-PSDB; ADO63818.  
XX  
XX New nucleic acid construct comprising heparanase-derived polypeptide, useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's thrombasthenia, or Bernard-Soulier syndrome.  
XX  
PS Claim 9; SEQ ID NO 9; 128pp; English.  
XX  
XX The invention relates to nucleic acid constructs comprising a nucleic acid encoding a heparanase-derived protein which lacks heparanase endoglycosidase catalytic activity but which retains its cell-cell and cell-matrix adhesion properties. The constructs of the invention optionally further comprise operably linked regulatory elements. The invention also relates to the heparanase-derived proteins and host cells comprising the nucleic acid constructs of the invention. The heparanase-derived proteins are especially mutants of human heparanase in which the active site proton donor Glu225 and/or the active site nucleophile Glu343 are replaced with Ala (ADO63822-ADO63824), and the proteins may optionally further comprise an avian heparanase signal peptide (ADO63825-

CC ADO63827). The heparanase-derived protein, nucleic acid construct and host cells are useful in preparing a tissue sealant composition for sealing injuries, reducing the loss of blood, accelerating the healing and homeostasis of an injury, accelerating blood vessel endothelium formation or the endothelialisation of vascular grafts, accelerating the adhesive activity of mammalian cells, and accelerating the adhesion and aggregation of platelets. They may also be use in the treatment of disorders associated with adhesion deficiency such as LAD (leukocyte adhesion deficiency), Glanzmann's thrombasthenia (defective platelet function), or Bernard-Soulier syndrome (deficient platelet adhesion). The cells of the invention may additionally be to screen for modulators of cell-cell and cell-matrix adhesion, and to prepare an implantable synthetic vascular graft comprising a tube made of a biocompatible material lined with the cells. The present sequence represents the human heparanase double mutant E225A/E343A.  
XX  
SQ Sequence 543 AA;  
Query Match 100.0%; Score 104; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 6.9e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFLIIDPK 19  
Db 89 PAYLRFGGTKTDFLIIDPK 107

RESULT 47  
ADO63823  
ID ADO63823 standard; protein; 543 AA.  
XX  
AC ADO63823;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE Human heparanase mutant E343A, SEQ ID:8.  
XX  
KW Human; heparanase; heparanase-derived protein; heparanase mutant;  
KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;  
KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;  
KW vascular graft; platelet adhesion; platelet aggregation;  
KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
KW vulnerary; mutant; mutein.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Active-site 225 /note= "Active site proton donor"  
FT Misc-difference 343 /note= "Ala replaces wild-type Glu (active site nucleophile)"  
XX  
PN WO2004048558-A2.  
XX  
PD 10-JUN-2004.  
XX  
PF 24-NOV-2003; 2003WO-IL000989.  
XX  
PR 24-NOV-2002; 2002IL-00153059.  
XX  
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
XX  
PI Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
XX  
DR WPI: 2004-450373/42.  
DR N-PSDB; ADO63817.  
XX  
XX New nucleic acid construct comprising heparanase-derived polypeptide, useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's thrombasthenia, or Bernard-Soulier syndrome.

XX Claim 9; SEQ ID NO 8; 128pp; English.

PS The invention relates to nucleic acid constructs comprising a nucleic

XX acid encoding a heparanase-derived protein which lacks heparanase

XX endoglycosidase catalytic activity but which retains its cell-cell and

CC cell-matrix adhesion properties. The constructs of the invention

CC optionally further comprise operably linked regulatory elements. The

CC invention also relates to the heparanase-derived proteins and host cells

CC comprising the nucleic acid constructs of the invention. The heparanase-

CC derived proteins are especially mutants of human heparanase in which the

CC active site proton donor Glu225 and/or the active site nucleophile Glu343

CC are replaced with Ala (ADO63822-ADO63824), and the proteins may

CC optionally further comprise an avian heparanase signal peptide (ADO63825-

CC ADO63827). The heparanase-derived protein, nucleic acid construct and

CC host cells are useful in preparing a tissue sealant composition for

CC sealing injuries, reducing the loss of blood, accelerating the healing

CC and homeostasis of an injury, accelerating blood vessel endothelium

CC formation or the endothelialisation of vascular grafts, accelerating the

CC adhesive activity of mammalian cells, and accelerating the adhesion and

CC aggregation of platelets. They may also be used in the treatment of

CC disorders associated with adhesion deficiency such as LAD (leukocyte

CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet

CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The

CC cells of the invention may additionally be used to screen for modulators of

CC cell-cell and cell-matrix adhesion, and to prepare an implantable

CC synthetic vascular graft comprising a tube made of a biocompatible

CC material lined with the cells. The present sequence represents the human

CC heparanase mutant E343A.

XX Sequence 543 AA;

XX

Query Match 100.0%; Score 104; DB 8; Length 543;

Best Local Similarity 100.0%; Pred. NO. 6.9e-09;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19

DB 89 PAYLRFGGTKTDFLIFDPK 107

RESULT 48

ADO63822

ID ADO63832 standard; protein; 543 AA.

XX

AC ADO63832;

XX

DT 26-AUG-2004 (first entry)

XX

DE Human heparanase mutant E396A.

XX

KW Human; heparanase; heparanase-derived protein; heparanase mutant;

KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;

KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;

KW vascular graft; platelet adhesion; platelet aggregation;

KW adhesion disorder; LAD; leukocyte adhesion deficiency;

KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;

KW vulnery; mutant; mutein; enzyme.

XX

OS Homo sapiens.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Active-site 225 /note= "Active site proton donor"

FT Active-site 343 /note= "Active site nucleophile"

FT Misc-difference 396 /note= "Ala replaces wild-type Glu"

FT

XX WO2004048558-A2.

PN

XX 10-JUN-2004.

PD

XX 24-NOV-2003; 2003WO-IL000989.

PF

XX 24-NOV-2002; 2002IL-00153059.

PR

XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

PA

XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;

PI

XX WPI; 2004-450373/42.

DR

XX New nucleic acid construct comprising heparanase-derived polypeptide,

XX useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's

PT thrombasthenia, or Bernard-Soulier syndrome.

PT

XX Example 4; Page; 128pp; English.

PS

XX The invention relates to nucleic acid constructs comprising a nucleic

XX acid encoding a heparanase-derived protein which lacks heparanase

CC endoglycosidase catalytic activity but which retains its cell-cell and

CC cell-matrix adhesion properties. The constructs of the invention

CC optionally further comprise operably linked regulatory elements. The

CC invention also relates to the heparanase-derived proteins and host cells

CC comprising the nucleic acid constructs of the invention. The heparanase-

CC derived proteins are especially mutants of human heparanase in which the

CC active site proton donor Glu225 and/or the active site nucleophile Glu343

CC are replaced with Ala (ADO63822-ADO63824), and the proteins may

CC optionally further comprise an avian heparanase signal peptide (ADO63825-

CC ADO63827). The heparanase-derived protein, nucleic acid construct and

CC host cells are useful in preparing a tissue sealant composition for

CC sealing injuries, reducing the loss of blood, accelerating the healing

CC and homeostasis of an injury, accelerating blood vessel endothelium

CC formation or the endothelialisation of vascular grafts, accelerating the

CC adhesive activity of mammalian cells, and accelerating the adhesion and

CC aggregation of platelets. They may also be used in the treatment of

CC disorders associated with adhesion deficiency such as LAD (leukocyte

CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet

CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The

CC cells of the invention may additionally be used to screen for modulators of

CC cell-cell and cell-matrix adhesion, and to prepare an implantable

CC synthetic vascular graft comprising a tube made of a biocompatible

CC material lined with the cells. The present sequence represents the human

CC heparanase mutant E343A.

XX Sequence 543 AA;

XX

Query Match 100.0%; Score 104; DB 8; Length 543;

Best Local Similarity 100.0%; Pred. NO. 6.9e-09;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIFDPK 19

DB 89 PAYLRFGGTKTDFLIFDPK 107

RESULT 49

ADO63822

ID ADO63822 standard; protein; 543 AA.

XX

AC ADO63822;

XX

DT 26-AUG-2004 (first entry)

XX

DE Human heparanase mutant E225A, SEQ ID:7.

XX

KW Human; heparanase; heparanase-derived protein; heparanase mutant;

KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;

KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;

KW vascular graft; platelet adhesion; platelet aggregation;

KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
XX vulnary; mutant; mutein.  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 225  
FT /note= "Ala replaces wild-type Glu (active site proton  
FT donor)"  
FT Active-site 343  
FT /note= "Active site nucleophile"  
XX WO2004048558-A2.  
XX  
XX 10-JUN-2004.  
XX  
XX 24-NOV-2003; 2003WO-IL000989.  
XX  
XX 24-NOV-2002; 2002IL-00153059.  
XX  
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
XX  
XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
XX  
XX WPI: 2004-450373/42.  
XX N-PSDB; ADO63816.  
XX  
XX New nucleic acid construct comprising heparanase-derived polypeptide,  
PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
PT thrombasthenia, or Bernard-Soulier syndrome.  
XX  
XX Claim 9; SEQ ID NO 7; 128pp; English.  
PS  
XX The invention relates to nucleic acid constructs comprising a nucleic  
CC acid encoding a heparanase-derived protein which lacks heparanase  
CC endoglycosidase catalytic activity but which retains its cell-cell and  
CC cell-matrix adhesion properties. The constructs of the invention  
CC optionally further comprise operably linked regulatory elements. The  
CC invention also relates to the heparanase-derived proteins and host cells  
CC comprising the nucleic acid constructs of the invention. The heparanase-  
CC derived proteins are especially mutants of human heparanase in which the  
CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
CC host cells are useful in preparing a tissue sealant composition for  
CC sealing injuries, reducing the loss of blood, accelerating the healing  
CC and homeostasis of an injury, accelerating blood vessel endothelium  
CC formation or the endothelialisation of vascular grafts, accelerating the  
CC adhesive activity of mammalian cells, and accelerating the adhesion and  
CC aggregation of platelets. They may also be use in the treatment of  
CC disorders associated with adhesion deficiency such as LAD (leukocyte  
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
CC cells of the invention may additionally be used to screen for modulators of  
CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
CC synthetic vascular graft comprising a tube made of a biocompatible  
CC material lined with the cells. The present sequence represents the human  
XX heparanase mutant E225A.  
XX  
SQ Sequence 543 AA;  
Query Match 100.0%; Score 104; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 6.9e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 PAYLRFGGTKTDFLIFDPK 19  
Db 89 PAYLRFGGTKTDFLIFDPK 107

RESULT 50  
ADQ80372  
ID ADQ80372 standard; protein; 543 AA.  
XX  
AC ADQ80372;  
XX  
DT 21-OCT-2004 (first entry)  
XX  
DE Heparanase protein.  
XX  
XX cytosatic; epidermal growth factor receptor modulator; identification;  
KW therapeutic response; cancer; EGFR; biomarker.  
KW  
XX Homo sapiens.  
XX  
XX WO2004063709-A2.  
XX  
XX 29-JUL-2004.  
XX  
XX 08-JAN-2004; 2004WO-US000368.  
XX  
XX 08-JAN-2003; 2003US-0438735P.  
XX  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX  
XX Amler LC, Januario T;  
XX  
XX WPI: 2004-544114/52.  
XX N-PSDB; ADQ80253.  
XX  
XX Identifying a mammal that will respond therapeutically to a method of  
PT treating cancer comprises comparing the level of a biomarker in a mammal  
PT before and after exposure to an epidermal growth factor receptor (EGFR)  
PT modulator.  
XX  
XX Disclosure; SEQ ID NO 144; 520pp; English.  
PS  
XX The invention relates to a method of identifying a mammal that will  
CC respond therapeutically to a method of treating cancer by administering  
CC an epidermal growth factor receptor (EGFR) modulator by comparing the  
CC level of a biomarker in a mammal before and after exposure to an EGFR  
CC modulator. The method comprises: (a) measuring, in the mammal, the level  
CC of at least one biomarker identified in the specification; (b) exposing  
CC the mammal to the EGFR modulator; and (c) measuring in the mammal the  
CC level of the biomarker where a difference in the level in step (c)  
CC compared to step (a) indicates that the mammal will respond  
CC therapeutically to the method of treating cancer. The method and  
CC biomarkers are useful for identifying a mammal that will respond  
CC therapeutically to a method of treating cancer by administering an  
CC epidermal growth factor receptor (EGFR) modulator. This sequence  
CC corresponds to one of the biomarkers whose levels of expression is  
CC measured in the method of the invention.  
XX  
SQ Sequence 543 AA;  
Query Match 100.0%; Score 104; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 6.9e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 PAYLRFGGTKTDFLIFDPK 19  
Db 89 PAYLRFGGTKTDFLIFDPK 107

Search completed: June 5, 2006, 12:41:31  
Job time : 162.219 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:43:17 ; Search time 21.0822 Seconds  
(without alignments)  
86.714 Million cell updates/sec

Title: US-10-645-659A-7

Perfect score: 104

Sequence: 1 PAYLRFGGTKTDFLIFDPK 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

1: PIR.80.\*

2: pir1.\*

3: pir2.\*

4: pir3.\*

5: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66	63.5	480	2	JC7506
2	49	47.1	588	2	D70394
3	48	46.2	264	1	CRH07
4	48	46.2	265	2	C86516
5	48	46.2	265	2	D72105
6	47	45.2	176	2	E72351
7	46	44.2	405	2	E30986
8	46	44.2	405	2	H85831
9	46	44.2	405	2	G84971
10	45	43.3	241	2	S77596
11	45	43.3	246	1	B55582
12	44	42.3	284	2	F84363
13	44	42.3	762	2	D87403
14	44	42.3	997	2	JC6067
15	44	42.3	998	2	A36368
16	43	41.3	242	1	S49346
17	43	41.3	244	1	B47468
18	43	41.3	253	2	B71346
19	43	41.3	1293	2	T01512
20	42.5	40.9	175	2	T21297
21	42	40.4	219	2	T26632
22	42	40.4	240	2	A69191
23	42	40.4	241	2	H81327
24	42	40.4	687	2	H72485
25	42	40.4	742	2	A49672
26	42	40.4	772	2	A55004
27	42	40.4	857	2	JC4169
28	41	39.4	144	2	T24474
29	41	39.4	227	2	S76494

30	41	39.4	262	1	S76610	hypothetical prote
31	41	39.4	444	2	A72209	hypothetical prote
32	41	39.4	471	2	A86303	hypothetical prote
33	41	39.4	660	2	A70362	N-methylhydantoina
34	41	39.4	785	2	B72608	probable hyua APE1
35	41	39.4	5069	2	T17464	rifamycin polyketi
36	40.5	38.9	164	2	B44827	FMRFamide-like pep
37	40	38.5	99	2	G98278	hypothetical prote
38	40	38.5	141	2	T51959	hypothetical prote
39	40	38.5	178	2	AD1156	hypothetical prote
40	40	38.5	221	2	AB2204	lipote-protein li
41	40	38.5	243	2	AG3447	cytochrome-c oxida
42	40	38.5	286	2	H97341	dihydrodipicolinat
43	40	38.5	310	2	T35417	probable beta-lact
44	40	38.5	338	2	D89102	protein F2555.2 li
45	40	38.5	346	2	AC0382	probable Pyridoxal
46	40	38.5	349	2	T03500	hypothetical prote
47	40	38.5	381	1	A47327	selenoprotein P pr
48	40	38.5	396	2	S56954	protein YBR162c ho
49	40	38.5	419	2	T20514	hypothetical prote
50	40	38.5	433	2	G71657	folypolyglutamate
51	40	38.5	438	2	B97712	hypothetical prote
52	40	38.5	528	2	S38242	hypothetical prote
53	40	38.5	576	2	T24193	hypothetical prote
54	40	38.5	585	2	T18885	hypothetical prote
55	40	38.5	742	2	A49340	alcohol dehydrogen
56	40	38.5	742	2	JS0326	alcohol dehydrogen
57	40	38.5	762	2	T00410	protein kinase hom
58	40	38.5	967	2	H86334	T20H2.10 protein -
59	39.5	38.0	223	2	B81378	two-component regu
60	39.5	38.0	246	2	T44624	cinI protein [lmpo
61	39.5	38.0	289	2	T35113	diaminopimelate ep
62	39.5	38.0	481	2	B80886	protein Tlfi.2 li
63	39	37.5	143	2	T24261	hypothetical prote
64	39	37.5	158	2	AE0444	hypothetical prote
65	39	37.5	186	2	C87500	conserved hypotet
66	39	37.5	192	2	A64098	molybdopterin-guan
67	39	37.5	243	1	S39989	cytochrome-c oxida
68	39	37.5	243	2	AG2765	cytochrome C oxida
69	39	37.5	243	2	D97546	cytochrome-c oxida
70	39	37.5	243	2	C95313	cytochrome-c oxida
71	39	37.5	243	2	H95344	FixO1 c-type cytoc
72	39	37.5	264	2	F83351	conserved hypotet
73	39	37.5	296	2	A64035	hypothetical prote
74	39	37.5	301	2	B71246	hypothetical prote
75	39	37.5	319	2	T45980	hypothetical prote
76	39	37.5	324	2	H81513	probable transcrip
77	39	37.5	329	2	H71192	hypothetical prote
78	39	37.5	339	2	A60918	FMRFamide precurs
79	39	37.5	380	2	T10442	selenoprotein P pr
80	39	37.5	385	1	OMRTSP	selenoprotein P pr
81	39	37.5	404	2	S57178	2-nitropropane dio
82	39	37.5	422	2	C70518	probable nant prot
83	39	37.5	425	2	AH1041	Vi polysaccharide
84	39	37.5	425	2	B36892	Vi polysaccharide
85	39	37.5	433	2	S75948	hypothetical prote
86	39	37.5	454	2	A43501	sucrose-6-phosphat
87	39	37.5	517	2	C69550	conserved hypotet
88	39	37.5	580	2	A83874	carbon starvation-
89	39	37.5	581	2	T50841	phosphinositide-s
90	39	37.5	661	2	A32475	dnak-type molecula
91	39	37.5	661	2	T15513	heat shock 70K pro
92	39	37.5	661	2	AB3513	ATPase virB4 homol
93	39	37.5	849	2	D82785	mannosyltransferas
94	39	37.5	898	2	H87481	ribonuclease, Rne/
95	39	37.5	1068	2	T04112	pol. protein homolo
96	38.5	37.0	400	2	E75135	methylmalonyl-coa
97	38.5	37.0	400	2	H71073	probable methylmal
98	38.5	37.0	695	2	T40168	hypothetical prote
99	38.5	37.0	984	2	T48216	hypothetical prote
100	38.5	37.0	1545	2	T14288	DNA (cytosine-5-)-
101	38.5	37.0	1761	2	T14289	DNA (cytosine-5-)-
102	38	36.5	134	2	F36789	hypothetical prote

103	38	36.5	134	2	AC0079	hypothetical prote	176	37	35.6	330	2	T35886	hypothetical prote
104	38	36.5	198	2	S60969	probable membrane	177	37	35.6	336	2	B71121	hypothetical prote
105	38	36.5	217	2	F87372	peptide methionine	178	37	35.6	350	2	AE3022	threonine aldolase
106	38	36.5	303	2	T29321	hypothetical prote	179	37	35.6	350	2	E98262	hypothetical prote
107	38	36.5	313	2	H64144	hypothetical prote	180	37	35.6	356	2	B90271	aspartate aminotra
108	38	36.5	313	2	C85036	hypothetical prote	181	37	35.6	382	2	S51962	FUN49 protein - ye
109	38	36.5	338	2	E72334	conserved hypothet	182	37	35.6	389	2	T44248	sarcosine oxidase
110	38	36.5	341	1	DEK601	glycerolaldehyde-3-p	183	37	35.6	390	2	T39975	sarcosine oxidase
111	38	36.5	341	1	DEK604	glycerolaldehyde-3-p	184	37	35.6	402	2	D82980	hypothetical prote
112	38	36.5	346	2	G86305	SRG1 homolog limpo	185	37	35.6	432	2	C81308	UDP-N-acetylmurama
113	38	36.5	359	2	T59403	alpha-2,8-polysial	186	37	35.6	438	2	A70544	probable histidino
114	38	36.5	359	2	S52425	polysialyltransfer	187	37	35.6	439	2	G97159	contains cell adhe
115	38	36.5	359	2	JC4224	alpha-N-acetylneur	188	37	35.6	440	2	T11319	NADH2 dehydrogenas
116	38	36.5	371	2	T21153	hypothetical prote	189	37	35.6	453	2	C83722	cytochrome P450 BH
117	38	36.5	379	1	DWDXB9	2-hydroxyglutaryl-	190	37	35.6	470	2	G70564	probable membrane
118	38	36.5	409	2	C95041	hypothetical prote	191	37	35.6	473	2	A84979	sulfate adenyllylr
119	38	36.5	409	2	A86240	protein F20B24.10	192	37	35.6	516	2	AC1132	hydantoinase homol
120	38	36.5	414	2	A87165	L-lactate dehydrog	193	37	35.6	516	2	AC1132	hydantoinase homol
121	38	36.5	426	2	C97797	tetrahydrofolylpol	194	37	35.6	529	2	F69020	conserved hypothet
122	38	36.5	440	2	F83235	outer membrane por	195	37	35.6	537	2	B46535	interleukin 2 rece
123	38	36.5	440	2	S11793	porin P precursor,	196	37	35.6	565	2	T47423	hypothetical prote
124	38	36.5	440	2	D71715	hypothetical prote	197	37	35.6	565	2	T29813	hypothetical prote
125	38	36.5	463	2	H72365	heat shock protein	198	37	35.6	578	2	B89045	protein B0238.7 li
126	38	36.5	464	2	B35159	arylsulfatase (EC	199	37	35.6	607	2	A26168	rhophorin I precu
127	38	36.5	521	2	T45608	hypothetical prote	200	37	35.6	607	2	S60905	hypothetical prote
128	38	36.5	553	2	S75892	probable glycerol-	201	37	35.6	639	2	C42049	leishmanolysin (EC
129	38	36.5	577	2	T45548	arylsulfatase (EC	202	37	35.6	644	2	E90367	n-methyl hydantoin
130	38	36.5	608	2	T32708	hypothetical prote	203	37	35.6	645	2	T49702	related to DOS1 pr
131	38	36.5	634	2	T25425	hypothetical prote	204	37	35.6	646	1	S19916	leishmanolysin (EC
132	38	36.5	725	2	D87002	probable cation tr	205	37	35.6	649	2	T33431	hypothetical prote
133	38	36.5	728	2	S55601	hypothetical prote	206	37	35.6	680	2	D64420	N-methylhydantoina
134	38	36.5	738	2	S14270	alcohol dehydrogen	207	37	35.6	684	2	T45534	agaf protein limpo
135	38	36.5	741	2	T12706	NADH2 dehydrogenas	208	37	35.6	787	2	C48998	hypothetical prote
136	38	36.5	834	2	S66498	M-sema F protein p	209	37	35.6	853	2	AC2079	ferrichrome-iron r
137	38	36.5	924	2	C66725	chromosome segrega	210	37	35.6	879	2	S49910	chloroplast outer
138	38	36.5	976	2	PC4208	valine-tRNA ligase	211	37	35.6	884	2	S66308	nitrate reductase
139	38	36.5	1048	2	T31653	hypothetical prote	212	37	35.6	891	2	AC3384	ribonuclease E / z
140	38	36.5	1163	2	T24855	hypothetical prote	213	37	35.6	936	2	D97630	hypothetical prote
141	38	36.5	1175	2	S39951	chitin synthase (E	214	37	35.6	1025	2	G81722	polymorphic membra
142	38	36.5	1205	2	AH2486	hypothetical prote	215	37	35.6	1289	2	S69689	hypothetical prote
143	38	36.5	1268	2	B36502	insulin receptor-r	216	37	35.6	1873	2	A30063	dihydropyridine re
144	38	36.5	1339	1	S20052	DNA-directed DNA p	217	37	35.6	2332	1	GNMYF	genome polyprotein
145	38	36.5	1374	2	T16129	probable polyketid	218	37	35.6	3083	2	AH2493	hypothetical prote
146	38	36.5	2100	2	T03223	probable multi-dom	219	36.5	35.1	91	2	S25462	Ig kappa chain v r
147	38	36.5	2240	2	T37057	hypothetical prote	220	36.5	35.1	95	2	S25177	Ig kappa chain v r
148	37.5	36.1	214	2	A87276	hypothetical prote	221	36.5	35.1	111	2	E53285	Ig kappa chain v a
149	37.5	36.1	347	2	A34616	FMRfamide polyprot	222	36.5	35.1	111	2	A33936	Ig kappa chain v r
150	37.5	36.1	362	2	G71167	hypothetical prote	223	36.5	35.1	131	1	KVNSM6	Ig kappa chain v r
151	37.5	36.1	375	2	JC6169	nuclear IIM intera	224	36.5	35.1	143	2	S77214	Ig kappa chain pre
152	37.5	36.1	386	2	T04914	hypothetical prote	225	36.5	35.1	269	2	D90141	hypothetical prote
153	37.5	36.1	430	2	T28143	tapasin 1 homolog,	226	36.5	35.1	280	2	T35432	probable aldehyde
154	37.5	36.1	566	2	T33042	hypothetical prote	227	36.5	35.1	313	2	B97425	flaC protein (X964
155	37.5	36.1	705	2	T51034	hypothetical prote	228	36.5	35.1	313	2	A12642	flagellin importe
156	37.5	36.1	1213	2	T51032	hypothetical prote	229	36.5	35.1	315	2	T35804	probable aldehyde
157	37.5	36.1	1456	2	T15961	hypothetical prote	230	36.5	35.1	448	2	F83628	beta-alanine-pyruv
158	37.5	36.1	1559	2	T07757	probable DNA (cyto	231	36.5	35.1	449	1	A42800	beta-alanine-pyruv
159	37.5	36.1	1699	2	T31340	voltage-gated sodi	232	36.5	35.1	464	2	H70169	phosphogluconate d
160	37.5	36.1	4588	2	T28667	dyein beta heavy	233	36.5	35.1	482	2	D83379	probable aldehyde
161	37	35.6	99	2	S28701	hypothetical prote	234	36.5	35.1	483	2	H98159	hypothetical prote
162	37	35.6	135	2	T21607	FMRfamide-like pep	235	36.5	35.1	483	2	A13127	vanillin: NAD oxia
163	37	35.6	198	2	T48931	hypothetical prote	236	36.5	35.1	545	1	S08020	RNA-directed RNA p
164	37	35.6	202	2	A64480	hypothetical prote	237	36.5	35.1	600	2	B84948	NADH2 dehydrogenas
165	37	35.6	224	2	G82913	conserved hypothet	238	36.5	35.1	653	2	F70383	organic solvent to
166	37	35.6	228	2	C69859	two-component resp	239	36.5	35.1	1400	2	A81672	secDF protein, pro
167	37	35.6	254	2	B84901	hypothetical prote	240	36.5	35.1	1554	2	T06370	probable DNA (cyto
168	37	35.6	275	2	E82554	extragenic suppress	241	36	34.6	70	2	S01213	NADH2 dehydrogenas
169	37	35.6	292	2	B82411	transcription regu	242	36	34.6	81	2	C86710	hypothetical prote
170	37	35.6	300	2	A92968	hypothetical prote	243	36	34.6	105	2	AC0277	probable phage pro
171	37	35.6	300	2	E98314	probable permease	244	36	34.6	107	2	A75089	hypothetical prote
172	37	35.6	315	2	T26673	hypothetical prote	245	36	34.6	115	2	AF1366	transcription regu
173	37	35.6	326	2	F82334	hflC protein VC035	246	36	34.6	115	2	AG1735	transcription regu
174	37	35.6	328	2	T16065	iron-sulfur cofact	247	36	34.6	126	2	T16727	hypothetical prote
175	37	35.6	330	2	A70422	hypothetical prote	248	36	34.6	137	2	T03491	conserved hypothet

249	36	34.6	143	2	A83328	probable ring-clea	322	36	34.6	634	2	S32349	probable SNF2-type
250	36	34.6	155	2	C97885	hypothetical prote	323	36	34.6	640	2	F90587	lipoprotein (impor
251	36	34.6	156	2	JQ0351	heat shock protein	324	36	34.6	650	2	A90473	n-methylhydantoina
252	36	34.6	156	2	T42664	heat shock protein	325	36	34.6	664	2	T39959	probable urea acti
253	36	34.6	157	2	S06074	heat shock protein	326	36	34.6	666	2	B70320	nitrate reductase
254	36	34.6	157	2	T14381	heat-shock protein	327	36	34.6	686	2	T10684	hypothetical prote
255	36	34.6	173	1	RUPSE0	rubredoxin II - Ps	328	36	34.6	722	1	VCPVCN	coat protein VP1 -
256	36	34.6	206	2	C83026	conserved hypotet	329	36	34.6	722	1	VCPVME	coat protein VP1 -
257	36	34.6	224	2	F64339	hypothetical prote	330	36	34.6	727	1	VCPVIF	coat protein VP1 -
258	36	34.6	229	2	I50106	MHC class I histoc	331	36	34.6	727	1	VCPVFP	coat protein VP1 -
259	36	34.6	239	2	B95374	hypothetical prote	332	36	34.6	735	2	I48101	ADAM 6 protein pre
260	36	34.6	244	2	S74911	hypothetical prote	333	36	34.6	737	1	VCPVCD	coat protein VP1 -
261	36	34.6	247	2	T11315	ATP synthase chain	334	36	34.6	741	2	I48694	probable transcrip
262	36	34.6	256	1	H70023	N-acetyl-glucosami	335	36	34.6	747	2	T34329	hypothetical prote
263	36	34.6	260	2	T08463	carbonate dehydrat	336	36	34.6	748	1	VCPVCP	coat protein VP1 -
264	36	34.6	262	2	S06060	gene NDI intron 4	337	36	34.6	749	2	GA4165	ATP-dependent heli
265	36	34.6	266	2	C95072	hypothetical prote	338	36	34.6	812	2	AC2349	hypothetical prote
266	36	34.6	271	2	AF7940	hypothetical prote	339	36	34.6	819	2	T22152	hypothetical prote
267	36	34.6	280	1	RLBH	rRNA N-glycosidase	340	36	34.6	820	2	A6247	hypothetical prote
268	36	34.6	281	2	B38664	30K ribosome inact	341	36	34.6	837	2	T19271	hypothetical prote
269	36	34.6	295	2	J47474	NAD-dinitrogen-red	342	36	34.6	844	1	A28528	penicillin amidase
270	36	34.6	295	2	I39751	NAD-dinitrogen-red	343	36	34.6	846	1	PNECA	penicillin amidase
271	36	34.6	298	2	T17568	hydrolase homolog	344	36	34.6	858	2	T47223	replication licens
272	36	34.6	318	2	C70661	probable moew - My	345	36	34.6	862	1	S56766	replication licens
273	36	34.6	320	2	T09555	fibrillarlin - Arab	346	36	34.6	863	1	S64720	replication licens
274	36	34.6	328	2	D84073	tellurite resistan	347	36	34.6	863	1	S65954	replication licens
275	36	34.6	336	2	A96997	ferrichrome transp	348	36	34.6	871	2	T40845	dna ligase - fissi
276	36	34.6	342	2	S76447	hypothetical prote	349	36	34.6	873	2	H96503	protein F9C16.17 [
277	36	34.6	343	2	A53057	retinal-binding pr	350	36	34.6	877	2	H71647	alanine-tRNA ligas
278	36	34.6	351	2	AF0559	probable lyase STY	351	36	34.6	878	2	G97865	alanine-tRNA ligas
279	36	34.6	370	2	C70464	GTP-binding protei	352	36	34.6	917	2	F95884	probable sensory h
280	36	34.6	373	2	F83020	UDP-glucose-heptos	353	36	34.6	969	2	F71418	hypothetical prote
281	36	34.6	374	2	AC0972	lipopolysaccharide	354	36	34.6	1070	2	S19686	alpha-glucosidase
282	36	34.6	374	2	B42595	glucosyltransferas	355	36	34.6	1116	2	T38073	serine/threonine-p
283	36	34.6	374	2	B98192	glucosyltransferas	356	36	34.6	1255	2	T31065	diaphanous protein
284	36	34.6	374	2	C96039	hypothetical prote	357	36	34.6	1260	2	T37523	probable oxoprolin
285	36	34.6	389	2	D83139	hypothetical prote	358	36	34.6	1483	2	T19751	hypothetical prote
286	36	34.6	401	2	S53862	NADH2 dehydrogenas	359	36	34.6	1517	2	T38912	hypothetical integ
287	36	34.6	404	2	T44590	tylosin biosynthes	360	36	34.6	1618	2	S21424	nestin - human
288	36	34.6	409	2	T11901	NADH2 dehydrogenas	361	36	34.6	1755	2	F82618	chemotaxis-related
289	36	34.6	415	2	T23215	hypothetical prote	362	36	34.6	2139	2	A44467	voltage-dependent
290	36	34.6	416	2	F70593	hypothetical prote	363	36	34.6	2143	2	JH0427	voltage-dependent
291	36	34.6	418	2	D81992	NADH2 dehydrogenas	364	36	34.6	2163	2	S50675	pre-mRNA splicing
292	36	34.6	418	2	B81222	NADH2 dehydrogenase	365	36	34.6	2166	2	S11339	calcium channel pr
293	36	34.6	424	2	H96963	dihydroorotase [im	366	36	34.6	2171	2	S05054	calcium channel al
294	36	34.6	426	2	T51373	hypothetical prote	367	36	34.6	2220	2	A45290	calcium channel pr
295	36	34.6	428	2	D83861	hypothetical prote	368	36	34.6	3491	2	T43231	probable 6-deoxyer
296	36	34.6	432	2	C54088	trigger factor tig	369	36	34.6	3573	2	S23070	erythronolide synt
297	36	34.6	438	2	A72430	hypothetical prote	370	36	34.6	3871	2	T22812	hypothetical prote
298	36	34.6	445	1	S26209	histidinol dehydro	371	35.5	34.1	65	2	B75298	hypothetical prote
299	36	34.6	452	2	T25076	hypothetical prote	372	35.5	34.1	73	2	A13065	hypothetical prote
300	36	34.6	454	2	A69763	homoserine dehydro	373	35.5	34.1	73	2	G98220	hypothetical prote
301	36	34.6	456	2	D70772	hypothetical prote	374	35.5	34.1	194	2	D69486	probable ribosomal
302	36	34.6	459	2	B44498	radial spoke prote	375	35.5	34.1	233	2	A13590	DNA-directed DNA p
303	36	34.6	474	2	H90582	preprotein translo	376	35.5	34.1	294	2	G82972	conserved hypotet
304	36	34.6	479	2	S68598	sucrose-6-phosphat	377	35.5	34.1	357	2	F75066	hypothetical prote
305	36	34.6	492	2	F64464	sodium-dependent n	378	35.5	34.1	378	1	T03758	probable ferredoxi
306	36	34.6	507	1	A32966	cytochrome P450 4A	379	35.5	34.1	378	2	A47300	cell adhesion prot
307	36	34.6	507	2	T01555	hypothetical prote	380	35.5	34.1	383	2	F89916	hippurate hydrolas
308	36	34.6	514	2	T44976	hydantoïnase homol	381	35.5	34.1	383	2	T23057	hypothetical prote
309	36	34.6	519	2	J05315	dihydropyrimidin	382	35.5	34.1	387	2	S56411	hypothetical 45K p
310	36	34.6	522	2	S33029	hypothetical prote	383	35.5	34.1	414	2	I39840	hypothetical prote
311	36	34.6	522	2	T21591	hypothetical prote	384	35.5	34.1	423	2	B83106	hypothetical prote
312	36	34.6	531	2	C75418	ribonucleoprotein	385	35.5	34.1	424	2	B56144	aspartate carbamoy
313	36	34.6	556	2	I61581	transcription fact	386	35.5	34.1	455	2	A0163	glycylpeptide N-te
314	36	34.6	590	1	A45621	leishmanolysin (EC	387	35.5	34.1	484	2	AG3184	aldenhyde dehydrog
315	36	34.6	599	2	B42049	leishmanolysin (EC	388	35.5	34.1	484	2	A64251	glutamate-tRNA lig
316	36	34.6	599	2	A44951	leishmanolysin (EC	389	35.5	34.1	486	2	T26483	hypothetical prote
317	36	34.6	603	2	G84554	probable acyl-CoA	390	35.5	34.1	508	1	A30007	dolichyl-diphospho
318	36	34.6	605	2	A27274	ribothorin 1 precu	391	35.5	34.1	508	1	ISHUSS	protein disulfide-
319	36	34.6	614	2	F96791	hypothetical prote	392	35.5	34.1	508	1	ISKTSS	protein disulfide-
320	36	34.6	628	2	H89917	conserved hypotet	393	35.5	34.1	509	1	ISMSSS	protein disulfide-
321	36	34.6	634	2	I40217	glutaryl 7-ACA acy	394	35.5	34.1	510	1	ISBOSS	protein disulfide-

395	35.5	34.1	515	1	ISCH9S	protein disulfide-	458	35	33.7	378	2	S77445	3-amino-5-hydroxyb
396	35.5	34.1	598	2	AG0311	NADH2 dehydrogenas	459	35	33.7	388	2	RR0450	hypothetical prote
397	35.5	34.1	735	2	F81993	DNA helicase II (E	460	35	33.7	390	1	GBPT4	recombination/rep
398	35.5	34.1	963	2	T04002	hypothetical prote	471	35	33.7	393	2	B72204	maltose ABC transp
399	35.5	34.1	980	2	T38632	probable phosphati	472	35	33.7	394	2	AG3616	hypothetical prote
400	35.5	34.1	1749	2	S75138	hypothetical prote	473	35	33.7	401	1	B35177	chromate resistanc
401	35	33.7	30	2	E31461	T-cell receptor de	474	35	33.7	409	1	S75020	acetyl-CoA C-acety
402	35	33.7	147	2	F71089	Ni,Fe-Hydrogenase	475	35	33.7	411	2	G97802	tyrosine-tRNA liga
403	35	33.7	156	2	A69966	hypothetical prote	476	35	33.7	412	2	T45097	hypothetical prote
404	35	33.7	156	2	A32868	prolactin receptor	477	35	33.7	412	2	A41070	prolactin receptor
405	35	33.7	158	2	F90127	hypothetical prote	478	35	33.7	419	2	H90480	conserved hypothet
406	35	33.7	162	2	G70232	hypothetical prote	479	35	33.7	426	2	T16406	hypothetical prote
407	35	33.7	168	2	E87709	hypothetical prote	480	35	33.7	428	2	H37118	levansucrase limpo
408	35	33.7	173	1	NKVGCU	core protein p20 -	481	35	33.7	432	2	D97851	UDP-glucose 6-dehy
409	35	33.7	173	2	A26659	T-cell receptor ga	482	35	33.7	433	2	A70536	probable pepC prot
410	35	33.7	179	2	D72360	conserved hypothet	483	35	33.7	434	2	E71638	UDP-glucose 6-dehy
411	35	33.7	195	2	D69505	phosphatidylserine	484	35	33.7	437	2	T44520	lipopolysaccharide
412	35	33.7	215	2	AB3358	probable carnitine	485	35	33.7	437	2	T44509	Vi polysaccharide
413	35	33.7	217	2	T27524	hypothetical prote	486	35	33.7	442	2	T34714	hypothetical prote
414	35	33.7	218	2	D96012	hypothetical membr	487	35	33.7	454	2	JC7231	thermophilic desul
415	35	33.7	224	2	AF3382	alpha/beta hydrola	488	35	33.7	464	2	C84428	probable ribophori
416	35	33.7	228	2	F72520	hypothetical prote	489	35	33.7	467	2	E69601	ATP-dependent Clp
417	35	33.7	228	2	H69025	hypothetical prote	490	35	33.7	469	1	AJZROL	glutamate-ammonia
418	35	33.7	232	2	S77110	hypothetical prote	491	35	33.7	471	1	JC2310	dihydropyrimidinas
419	35	33.7	237	2	D75359	DNA polymerase-rel	492	35	33.7	479	2	T29720	hypothetical prote
420	35	33.7	248	2	G72211	conserved hypothet	493	35	33.7	491	2	A36036	cytochrome P450 2F
421	35	33.7	251	2	C82878	integrase-recombin	494	35	33.7	494	2	B69363	hydantoin utilizat
422	35	33.7	252	2	C91263	phosphonate metabo	495	35	33.7	496	2	AH3254	acetyl-CoA:acetoac
423	35	33.7	252	2	H86103	phosphonate metabo	496	35	33.7	512	2	A98352	hydantoinase homol
424	35	33.7	252	2	H35719	pHP protein - Esc	497	35	33.7	512	2	AE2930	hydantoinase A [im
425	35	33.7	254	2	C84315	hypothetical prote	498	35	33.7	513	2	F83900	long-chain acyl-Co
426	35	33.7	255	2	H70467	HMP-p kinase - Aqu	499	35	33.7	517	2	D84421	probable amino aci
427	35	33.7	259	1	CRBO2	carbonate dehydrat	500	35	33.7	528	2	I48253	beta-N-acetylhexos
428	35	33.7	262	2	E96937	chemotaxis motilit	501	35	33.7	529	1	AQHUBA	beta-N-acetylhexos
429	35	33.7	270	2	B86925	conserved hypothet	502	35	33.7	548	2	D69187	probable acid-CoA
430	35	33.7	270	2	E70669	probable glycosylt	503	35	33.7	552	2	A67556	hypothetical prote
431	35	33.7	274	2	S33440	T-cell receptor ga	504	35	33.7	555	2	E87003	probable DNA methy
432	35	33.7	275	2	G82671	protoporphyrinogen	505	35	33.7	569	2	D9812	regulator of patho
433	35	33.7	279	1	C42646	dihydropterolate sy	506	35	33.7	570	2	S09812	hypothetical prote
434	35	33.7	279	1	SYECOG	dihydropterolate sy	507	35	33.7	583	2	T39112	probable amidase -
435	35	33.7	279	2	S07654	dihydropterolate sy	508	35	33.7	608	2	I3269	prolactin receptor
436	35	33.7	279	2	T45123	dihydropterolate sy	509	35	33.7	610	2	A34631	lactogen receptor
437	35	33.7	280	2	JC5848	protein synthesis	510	35	33.7	610	2	A36116	prolactin receptor
438	35	33.7	280	2	A95936	probable alcohol s	511	35	33.7	615	2	D96499	probable UDP-gluc
439	35	33.7	283	2	S10928	dihydropterolate sy	512	35	33.7	619	2	E83635	hypothetical prote
440	35	33.7	292	2	S74614	hypothetical prote	513	35	33.7	619	2	H84416	hypothetical prote
441	35	33.7	292	2	I77525	prolactin receptor	514	35	33.7	620	2	T10423	estrogen receptor
442	35	33.7	302	2	AH0550	hypothetical ROK-f	515	35	33.7	633	2	B70946	NADH2 dehydrogenas
443	35	33.7	303	2	I77524	site-specific DNA-	516	35	33.7	650	2	S22835	alpha-agglutinin -
444	35	33.7	304	2	C64109	hypothetical prote	517	35	33.7	657	2	T34037	heat shock 70K pro
445	35	33.7	305	2	A83340	fructokinase (EC 2	518	35	33.7	659	2	S10228	paraspinal crystal
446	35	33.7	307	1	JQ0782	stage III sporulat	519	35	33.7	661	2	I56258	Rp105 - mouse
447	35	33.7	307	2	S16622	prolactin receptor	520	35	33.7	666	2	T10567	probable serine/th
448	35	33.7	310	2	A29884	hypothetical prote	521	35	33.7	666	2	T05432	hypothetical prote
449	35	33.7	315	2	AD3127	probable transcrip	522	35	33.7	677	2	T08943	hypothetical prote
450	35	33.7	315	2	E83369	probable transcrip	523	35	33.7	681	1	H82059	2',3'-cyclic-nucle
451	35	33.7	317	2	B83158	probable transcrip	524	35	33.7	682	2	T12968	hypothetical prote
452	35	33.7	318	2	S37225	ribose-phosphate d	525	35	33.7	683	1	S69780	outer membrane pro
453	35	33.7	318	2	F87529	sugar isomerase, K	526	35	33.7	690	2	H69268	copper-transferrin
454	35	33.7	319	2	E97792	kpsF protein [limp	527	35	33.7	720	2	T15756	hypothetical prote
455	35	33.7	319	2	E98160	hypothetical prote	528	35	33.7	735	2	S46830	urea transport pro
456	35	33.7	322	2	B87419	3-oxoacyl-(acyl-ca	529	35	33.7	771	2	H72410	hypothetical prote
457	35	33.7	329	2	T10683	hypothetical prote	530	35	33.7	777	2	F82560	hypothetical prote
458	35	33.7	332	2	AH1562	mylodopterin bios	531	35	33.7	780	2	T09485	hypothetical prote
459	35	33.7	332	2	AD3340	cobC protein [limp	532	35	33.7	784	2	A86676	cold-induced prote
460	35	33.7	336	2	AT0183	galactoside permea	533	35	33.7	832	2	T23693	hypothetical prote
461	35	33.7	338	2	E97086	anaerobic sulfite	534	35	33.7	853	2	S59315	hypothetical prote
462	35	33.7	339	2	T28019	hypothetical prote	535	35	33.7	855	2	C82983	hypothetical prote
463	35	33.7	342	2	A46396	ets-related protei	536	35	33.7	879	2	T49796	probable sepB prot
464	35	33.7	346	3	A95361	probable transcrip	537	35	33.7	902	2	T41051	beta transducin -
465	35	33.7	356	2	JQ1146	C-5 sterol desatur	538	35	33.7	935	2	E96806	hypothetical prote
466	35	33.7	373	1	H69281	conserved hypothet	539	35	33.7	973	2	A97522	ribonuclease E, RN
467	35	33.7	374	2	T27420	hypothetical prote	540	35	33.7	977	2	AC2741	ribonuclease E [im



541	35	33.7	982	2	A97210	beta galactosidase	614	34	32.7	121	2	T42138	type II secretion
542	35	33.7	1023	2	B38932	phospholipase C (E	615	34	32.7	125	2	A70356	hypothetical prote
543	35	33.7	1060	2	E83547	proline dehydrogen	616	34	32.7	133	2	AF0391	curlin genes regul
544	35	33.7	1108	2	I39385	guanylate cyclase	617	34	32.7	134	2	AG2526	conserved hypotnet
545	35	33.7	1132	2	H82887	hypothetical prote	618	34	32.7	134	2	H98355	hypothetical prote
546	35	33.7	1162	2	T49191	hypothetical prote	619	34	32.7	141	1	HADD	hemoglobin alpha c
547	35	33.7	1163	2	JR0366	tight junction pro	620	34	32.7	141	2	T20663	hypothetical prote
548	35	33.7	1169	2	T18423	hypothetical prote	621	34	32.7	146	2	S36323	T-cell receptor de
549	35	33.7	1171	2	A80130	DNA-directed DNA p	622	34	32.7	147	2	D75132	probable hydrogena
550	35	33.7	1172	2	T49330	cytokinesis inhibi	623	34	32.7	158	2	G75542	probable RNA methy
551	35	33.7	1176	1	A40447	phospholipase C (E	624	34	32.7	163	2	H69990	sugar transport pr
552	35	33.7	1179	2	T04488	DNA topoisomerase	625	34	32.7	167	2	AE2342	hypothetical prote
553	35	33.7	1186	2	AD1300	Smc protein essent	626	34	32.7	167	2	E71943	hypothetical prote
554	35	33.7	1186	2	AD1672	Smc protein essent	627	34	32.7	170	2	T42413	FMRFamide-like pep
555	35	33.7	1205	2	B95845	conserved hypotnet	628	34	32.7	175	2	AF1512	hypothetical prote
556	35	33.7	1208	2	AI3584	N-methylhydantoina	629	34	32.7	179	2	A95245	hypothetical prote
557	35	33.7	1217	2	T00270	hypothetical prote	630	34	32.7	183	2	G96018	probable methylate
558	35	33.7	1275	2	T18556	O-antigen biosynth	631	34	32.7	199	2	A82360	conserved hypotnet
559	35	33.7	1321	2	A60165	sodium channel pro	632	34	32.7	207	2	G95238	conserved hypotnet
560	35	33.7	1341	2	S50366	probable membrane	633	34	32.7	207	2	H98102	hypothetical prote
561	35	33.7	1386	2	T49316	profilaggrin relat	634	34	32.7	208	2	T20554	hypothetical prote
562	35	33.7	1413	2	D84481	probable retroelem	635	34	32.7	213	2	I39900	conserved hypotnet
563	35	33.7	1498	2	AG1439	B. subtilis Yuka p	636	34	32.7	214	2	C82255	adenylate kinase V
564	35	33.7	1832	2	T31113	mucin-like glycopr	637	34	32.7	216	1	BVECKB	alkB protein - Esc
565	35	33.7	2126	2	H70621	probable polyketid	638	34	32.7	216	2	E91016	alkylated DNA repa
566	35	33.7	2179	1	GNNYH4	genome polyprotein	639	34	32.7	216	2	G85860	DNA repair system
567	35	33.7	2262	2	T30890	calcium channel al	640	34	32.7	219	2	AG0872	probable membrane
568	35	33.7	2337	2	T40577	probable phosphati	641	34	32.7	219	2	C61074	probable oxidoredu
569	35	33.7	2493	2	S45734	probable membrane	642	34	32.7	219	2	C65074	probable oxidoredu
570	35	33.7	3169	2	T00296	toxin B - Escheric	643	34	32.7	219	2	S59967	hemolysin III - Ba
571	34.5	33.2	142	2	S36310	T-cell receptor de	644	34	32.7	219	2	G85945	probable oxidoredu
572	34.5	33.2	238	1	TRWV5X	trypsin-like prote	645	34	32.7	222	2	B54898	STX protein - huma
573	34.5	33.2	255	2	AE0537	hypothetical prote	646	34	32.7	225	2	A35295	glutathione transf
574	34.5	33.2	281	2	T24658	hypothetical prote	647	34	32.7	228	2	B86836	hypothetical prote
575	34.5	33.2	290	2	A46353	ORF1 protein - coc	648	34	32.7	231	2	JO1247	hypothetical 25.7K
576	34.5	33.2	322	1	A70189	hypothetical prote	649	34	32.7	235	2	A97996	degenerate transpo
577	34.5	33.2	326	2	AF2024	hypothetical prote	650	34	32.7	241	2	AG2915	conserved hypotnet
578	34.5	33.2	334	2	AG2190	hypothetical prote	651	34	32.7	241	2	B97690	hypothetical prote
579	34.5	33.2	338	2	T23520	hypothetical prote	652	34	32.7	243	2	C95908	hypothetical prote
580	34.5	33.2	394	2	G86657	ABC transporter At	653	34	32.7	247	2	C87423	cytochrome c oxida
581	34.5	33.2	412	2	T40295	fructosyl amine -	654	34	32.7	253	2	H69046	hypothetical prote
582	34.5	33.2	450	2	G83740	hypothetical prote	655	34	32.7	256	1	A42768	homeotic protein g
583	34.5	33.2	466	2	AH1213	TN916 ORF21 homolo	656	34	32.7	256	2	AG2873	dehydrogenase Atu2
584	34.5	33.2	511	2	AB0858	hypothetical prote	657	34	32.7	256	2	H97649	probable gluconate
585	34.5	33.2	538	2	T06683	aldehyde dehydroge	658	34	32.7	256	2	AB1098	conserved hypotnet
586	34.5	33.2	570	2	T02676	hypothetical prote	659	34	32.7	257	2	AB1098	conserved hypotnet
587	34.5	33.2	600	1	D65000	NADH2 dehydrogenas	660	34	32.7	257	2	AI1460	conserved hypotnet
588	34.5	33.2	600	2	A00797	NADH2 dehydrogenas	661	34	32.7	259	1	CR8B2	carbonate dehydrat
589	34.5	33.2	600	2	C85869	NADH dehydrogenase	662	34	32.7	260	1	CRH2	hypothetical prote
590	34.5	33.2	600	2	B91025	NADH dehydrogenase	663	34	32.7	261	2	T22623	hypothetical prote
591	34.5	33.2	603	2	T02677	hypothetical prote	664	34	32.7	263	2	B82731	UDP-N-acetylglucos
592	34.5	33.2	605	1	A35459	glucose oxidase (E	665	34	32.7	266	2	B89886	glutamate racemase
593	34.5	33.2	611	2	T28171	hypothetical prote	666	34	32.7	268	2	G02133	holocytochrome-c s
594	34.5	33.2	612	2	A44857	acetylactate synth	667	34	32.7	268	2	T27242	hypothetical prote
595	34.5	33.2	635	1	A293358	cerevisin (EC 3.4.	668	34	32.7	270	2	T09514	5'-AMP-activated p
596	34.5	33.2	693	2	AF2357	hypothetical prote	669	34	32.7	276	2	D70191	hypothetical prote
597	34.5	33.2	705	1	TFCHE	ovotransferrin pre	670	34	32.7	280	2	E75216	hypothetical prote
598	34.5	33.2	856	2	A64699	hypothetical prote	671	34	32.7	280	2	C71453	hypothetical prote
599	34.5	33.2	967	2	C70831	probable mmpL4 pro	672	34	32.7	282	2	E72415	zinc ABC transport
600	34.5	33.2	968	2	F70746	probable mmpL2 pro	673	34	32.7	288	2	S03603	uncoupling protein
601	34.5	33.2	1744	1	CAHU	complement C4a pre	674	34	32.7	291	2	A90084	hypothetical prote
602	34	32.7	55	2	AE2103	hypothetical prote	675	34	32.7	294	2	C72689	probable diptine
603	34	32.7	61	2	B84091	hypothetical prote	676	34	32.7	299	2	AC2998	glutamyl-trNA synt
604	34	32.7	65	2	B38601	Ig kappa chain v r	677	34	32.7	299	2	F98285	HA-exporting ANPas
605	34	32.7	85	2	S25837	hypothetical prote	678	34	32.7	309	2	T12089	hypothetical prote
606	34	32.7	93	2	A38601	Ig kappa chain v r	679	34	32.7	309	2	T09408	beta-carotene hydr
607	34	32.7	94	2	S44785	C30C11.3 protein -	680	34	32.7	310	2	T09562	hypothetical prote
608	34	32.7	108	1	K1HURY	Ig kappa chain v-I	681	34	32.7	311	2	T15997	hypothetical prote
609	34	32.7	113	2	E44151	Ig lambda chain v	682	34	32.7	312	2	T02406	hypothetical prote
610	34	32.7	113	2	T00196	hypothetical prote	683	34	32.7	313	2	H70777	probable coob - My
611	34	32.7	116	2	T42412	FMRFamide-like pep	684	34	32.7	325	1	A48561	inner capsid prote
612	34	32.7	117	2	AH0623	probable secreted	685	34	32.7	335	2	G75040	probable aminopept
613	34	32.7	121	2	S26798	Ig heavy chain v r	686	34	32.7	335	2	B98142	hypothetical prote

687	34	32.7	340	2	G75531	cytochrome c famil	760	34	32.7	484	1	A35282	alpha-amylase (EC
688	34	32.7	343	2	H96984	endoglucanase, ami	761	34	32.7	486	1	S63251	homoserine O-acetyl
689	34	32.7	345	2	T16427	hypothetical prote	762	34	32.7	489	2	T09151	glutathione-disulf
690	34	32.7	347	2	B87274	hypothetical prote	763	34	32.7	499	2	B83493	conserved hypothet
691	34	32.7	356	2	A13145	GDP-mannose 4,6-de	764	34	32.7	504	2	A83286	acetyl-CoA C-acety
692	34	32.7	357	1	E64649	probable X-Pro dip	765	34	32.7	506	2	S63181	hypothetical prote
693	34	32.7	357	2	A71938	probable proline p	766	34	32.7	508	2	A82958	hypothetical prote
694	34	32.7	357	2	B98133	hypothetical prote	767	34	32.7	509	1	O4RTLO	laurate omega-hydr
695	34	32.7	357	2	A83154	conserved hypothet	768	34	32.7	513	2	G83375	probable glycogen
696	34	32.7	359	2	F71136	hypothetical prote	769	34	32.7	517	2	G89324	hypothetical prote
697	34	32.7	363	2	A82065	threonine synthase	770	34	32.7	518	2	A12929	hydantoinase [impo
698	34	32.7	373	2	T48381	hypothetical prote	771	34	32.7	518	2	E98352	hydantoinase (AP00
699	34	32.7	375	2	T02215	ferredoxin-NADP re	772	34	32.7	524	2	T09937	cytokinin oxidase
700	34	32.7	375	2	I39169	sialyltransferase	773	34	32.7	529	1	H70736	IMP dehydrogenase
701	34	32.7	375	2	I48686	N-glycan alpha 2,8	774	34	32.7	530	2	T43089	transfer complex p
702	34	32.7	375	2	A46727	sialyltransferase	775	34	32.7	533	2	D83885	hydantoinase BH188
703	34	32.7	378	2	T02977	ferredoxin-NADP re	776	34	32.7	542	2	T02379	hypothetical prote
704	34	32.7	380	2	A83834	transposase (02) B	777	34	32.7	544	2	C82900	probable ABC subst
705	34	32.7	380	2	AB1709	N-acetylglucosamin	778	34	32.7	550	1	VGBE18	glycoprotein E - h
706	34	32.7	382	2	B86430	hypothetical prote	779	34	32.7	555	2	D95139	DNA repair protein
707	34	32.7	383	1	VGBEXG	glycoprotein precu	780	34	32.7	555	2	C98007	DNA repair and gen
708	34	32.7	383	2	H95970	probable sugar upt	781	34	32.7	573	2	S68591	methanol dehydroge
709	34	32.7	391	2	T27918	hypothetical prote	782	34	32.7	578	2	T11659	hypothetical prote
710	34	32.7	393	2	T23602	hypothetical prote	783	34	32.7	580	2	S63391	probable membrane
711	34	32.7	394	2	G72212	exodeoxyribonuclea	784	34	32.7	581	2	C39768	cholinesterase (EC
712	34	32.7	395	2	B47071	nitrogenase cofact	785	34	32.7	600	2	A49230	dnak-type molecula
713	34	32.7	396	2	AF0176	conserved hypothet	786	34	32.7	602	1	ACHU	cholinesterase (EC
714	34	32.7	403	2	A28451	beta-galactoside a	787	34	32.7	602	2	A53583	geranylgeranyl-dip
715	34	32.7	404	2	B83769	transposase (02) B	788	34	32.7	608	2	F83397	probable peptidase
716	34	32.7	404	2	B83797	transposase (02) B	789	34	32.7	611	2	D70928	hypothetical prote
717	34	32.7	404	2	B83817	transposase (02) B	790	34	32.7	615	2	G82658	periplasmic glucan
718	34	32.7	404	2	F83860	transposase (02) B	791	34	32.7	627	2	T32958	hypothetical prote
719	34	32.7	404	2	E83862	transposase (02) B	792	34	32.7	637	2	JQ2039	polyprotein - pars
720	34	32.7	404	2	C83895	transposase (02) B	793	34	32.7	639	2	B86587	DNA helicase [impo
721	34	32.7	404	2	D83925	transposase (02) B	794	34	32.7	639	2	G72038	ATP-dependent heli
722	34	32.7	404	2	H83947	transposase (02) B	795	34	32.7	641	1	UVBPAL	DNA-packaging prot
723	34	32.7	404	2	A83958	transposase (02) B	796	34	32.7	641	2	JN0668	dnak-type molecula
724	34	32.7	404	2	E83959	transposase (02) B	797	34	32.7	641	2	H85689	terminase large su
725	34	32.7	404	2	A83980	transposase (02) B	798	34	32.7	641	2	F90832	terminase large su
726	34	32.7	404	2	C83981	transposase (02) B	799	34	32.7	641	2	C90901	probable terminase
727	34	32.7	404	2	B84015	transposase (02) B	800	34	32.7	650	2	S75072	probable phytoene
728	34	32.7	404	2	B84022	transposase (02) B	801	34	32.7	650	2	G83465	conserved hypothet
729	34	32.7	404	2	G84055	transposase (02) B	802	34	32.7	654	2	H86717	hypothetical prote
730	34	32.7	404	2	A84076	transposase (02) B	803	34	32.7	656	1	A34890	histidine decarbox
731	34	32.7	404	2	C84088	transposase (02) B	804	34	32.7	658	2	S39653	xcpQ protein - Pse
732	34	32.7	404	2	E84095	transposase (02) B	805	34	32.7	662	1	A49882	histidine decarbox
733	34	32.7	404	2	T27971	hypothetical prote	806	34	32.7	662	1	S12889	histidine decarbox
734	34	32.7	406	2	S68866	qsopA protein - Co	807	34	32.7	662	2	G95929	probable methylcro
735	34	32.7	420	2	A13614	hypothetical prote	808	34	32.7	662	2	E97973	hypothetical prote
736	34	32.7	422	2	C70370	dihydroorotase - A	809	34	32.7	685	2	T09159	RNA helicase prh75
737	34	32.7	425	2	C87563	homogentisate 1,2-	810	34	32.7	698	2	H95890	dnak protein (heat
738	34	32.7	427	2	D81784	probable integral	811	34	32.7	704	2	H95890	probable hydantoin
739	34	32.7	427	2	G81208	AmpG-related prote	812	34	32.7	704	2	AH1958	hypothetical prote
740	34	32.7	428	2	T05521	alpha-amylase (EC	813	34	32.7	741	2	T12762	NADH2 dehydrogenas
741	34	32.7	430	2	F87711	heat shock protein	814	34	32.7	751	1	SYBYMT	methionine-tRNA li
742	34	32.7	434	2	AB3508	heat shock protein	815	34	32.7	758	1	SUECCA	ATP-dependent Clp
743	34	32.7	437	2	T37459	homogentisate 1,2-	816	34	32.7	758	2	A10609	ATP-binding compo
744	34	32.7	448	2	C97757	hypothetical prote	817	34	32.7	758	2	H90749	ATP-binding compo
745	34	32.7	449	2	T19626	hypothetical prote	818	34	32.7	758	2	D85600	Shaw type potassi
746	34	32.7	450	2	B70318	chaperone HslU - A	819	34	32.7	769	2	I56546	Imbition protein
747	34	32.7	450	2	B97758	heat shock protein	820	34	32.7	773	2	T46188	glycoprotein X pre
748	34	32.7	450	2	B71688	heat shock protein	821	34	32.7	797	1	VGBEX1	GRP diphosphokinas
749	34	32.7	458	2	S44741	Co2C2_3 protein -	822	34	32.7	847	2	S70687	hypothetical prote
750	34	32.7	461	2	AE0311	NADH2 dehydrogenas	823	34	32.7	862	2	F75116	membrane glycoprot
751	34	32.7	465	2	D84427	hypothetical prote	824	34	32.7	866	2	T45462	membrane glycoprot
752	34	32.7	466	2	T28216	hypothetical prote	825	34	32.7	867	2	T14777	hypothetical prote
753	34	32.7	469	2	AC2794	glutamine syntheta	826	34	32.7	867	2	AT1878	replicative DNA he
754	34	32.7	469	2	B97573	glutamine syntheta	827	34	32.7	879	2	A11878	hypothetical prote
755	34	32.7	469	2	B75039	hypothetical prote	828	34	32.7	882	2	A39030	androgen-binding p
756	34	32.7	471	2	A10362	glutamate-tRNA lig	829	34	32.7	895	2	B64238	isoleucine-tRNA li
757	34	32.7	473	2	AF2433	aldehyde dehydroge	830	34	32.7	896	1	A35782	cytokine receptor
758	34	32.7	473	2	T04482	ribophorin I homol	831	34	32.7	905	2	G84582	hypothetical prote
759	34	32.7	483	2	T21327	hypothetical prote	832	34	32.7	914	2	E83901	hypothetical prote

833	34	32.7	934	2	B83789	hypothetical prote	906	33.5	32.2	428	2	B82248	long-chain fatty
834	34	32.7	935	2	A40694	cadherin-associate	907	33.5	32.2	431	2	A56822	synaptonemal comp
835	34	32.7	948	2	T47322	plasma membrane H+	908	33.5	32.2	439	2	T02453	hypothetical prote
836	34	32.7	949	2	G84486	probable plasma me	909	33.5	32.2	486	2	T18903	hypothetical prote
837	34	32.7	953	2	S55156	probable membrane	910	33.5	32.2	551	2	D64412	hypothetical prote
838	34	32.7	966	2	F84582	hypothetical prote	911	33.5	32.2	556	2	T39109	probable guanosine
839	34	32.7	981	2	T46330	hypothetical prote	912	33.5	32.2	593	2	D83316	NADH dehydrogenase
840	34	32.7	990	2	T05197	hypothetical prote	913	33.5	32.2	599	2	B72368	conserved hypotnet
841	34	32.7	1014	2	T13476	hypothetical prote	914	33.5	32.2	602	2	T47794	hypothetical prote
842	34	32.7	1017	2	T31354	hypothetical prote	915	33.5	32.2	632	2	T48316	hypothetical prote
843	34	32.7	1034	2	T30574	probable potassium	916	33.5	32.2	686	2	AI0020	probable membrane
844	34	32.7	1034	2	JC5598	beta-galactosidase	917	33.5	32.2	794	2	B95942	probable aldehyde
845	34	32.7	1042	2	A57534	mucin - rat	918	33.5	32.2	811	2	T19974	hypothetical prote
846	34	32.7	1063	2	D83789	mucin SAC (clone L	919	33.5	32.2	816	2	S19139	hypothetical prote
847	34	32.7	1106	2	S38783	hypothetical prote	920	33.5	32.2	859	2	F69159	sucrose synthase (
848	34	32.7	1123	2	A72311	integrin alpha cha	921	33.5	32.2	859	2	F69159	protoporphyrin IX
849	34	32.7	1131	2	F88570	conserved hypotnet	922	33.5	32.2	939	2	T32521	hypothetical prote
850	34	32.7	1152	2	AE1852	protein K03H1.2 [i	923	33.5	32.2	1904	2	T13256	hypothetical prote
851	34	32.7	1165	1	S45879	hypothetical prote	924	33.5	32.2	2344	2	T41590	tail-host specific
852	34	32.7	1442	2	T42607	chitin synthase [E	925	33.5	32.2	2767	1	UIHU	probable sensor-li
853	34	32.7	1451	2	S41025	transcription acti	926	33	31.7	49	2	B82007	thyroglobulin prote
854	34	32.7	1587	2	AB2012	hypothetical prote	927	33	31.7	57	2	D81226	hypothetical prote
855	34	32.7	1621	2	A92255	hypothetical prote	928	33	31.7	61	2	H95874	hypothetical prote
856	34	32.7	1708	1	A43100	hypothetical prote	929	33	31.7	64	2	S09280	ferredoxin 2 [4Fe-4
857	34	32.7	1873	2	A55645	ataxia telangiecta	930	33	31.7	68	2	H98098	hypothetical prote
858	34	32.7	1879	2	T19481	calcium channel, v	931	33	31.7	74	2	T07559	hypothetical prote
859	34	32.7	1891	2	T43262	hypothetical prote	932	33	31.7	76	2	T28248	ORF MSV087 probabl
860	34	32.7	2476	2	T34022	calcium channel al	933	33	31.7	81	2	G85844	unknown protein en
861	34	32.7	4427	2	PN0637	zonahesin - pig	934	33	31.7	88	1	EDB251	immediate-early-5
862	34	32.7	5255	2	T31677	polyketide synthas	935	33	31.7	90	2	A69323	hypothetical prote
863	33.5	32.2	62	2	T36967	bacitracin synthet	936	33	31.7	99	2	S20342	hypothetical prote
864	33.5	32.2	62	2	AH0880	gene MHC DQ-alpha	937	33	31.7	108	2	S64602	calcium-binding pr
865	33.5	32.2	63	2	I36900	hypothetical prote	938	33	31.7	110	2	T10451	hypothetical prote
866	33.5	32.2	63	2	I61798	gene MHC DQ-alpha	939	33	31.7	120	2	S36306	PFPI protein - whi
867	33.5	32.2	63	2	I61801	gene MHC DQ-alpha	940	33	31.7	120	2	S57405	T-cell receptor de
868	33.5	32.2	63	2	I61787	gene MHC DQ-alpha	941	33	31.7	121	2	E64391	response regulator
869	33.5	32.2	63	2	I61789	gene MHC DQ-alpha	942	33	31.7	121	2	AI1646	hypothetical prote
870	33.5	32.2	63	2	H34513	gene MHC DQ-alpha	943	33	31.7	124	2	A11646	hypothetical prote
871	33.5	32.2	63	2	G34513	MHC class II histo	944	33	31.7	124	2	B86802	prophage p13 prote
872	33.5	32.2	63	2	B34513	MHC class II histo	945	33	31.7	133	2	S15472	hypothetical prote
873	33.5	32.2	63	2	C34514	MHC class II histo	946	33	31.7	133	2	D72110	hypothetical prote
874	33.5	32.2	63	2	C34513	MHC class II histo	947	33	31.7	133	2	E86513	hypothetical prote
875	33.5	32.2	87	2	C30575	MHC class II histo	948	33	31.7	138	2	JC4597	hypothetical prote
876	33.5	32.2	87	2	B27628	MHC class II histo	949	33	31.7	138	2	EG4597	signal transductio
877	33.5	32.2	142	2	S17462	glutathione transf	950	33	31.7	141	1	HA1G1	hemoglobin alpha-i
878	33.5	32.2	158	2	G95925	glutathione transf	951	33	31.7	141	1	HA1G1	hemoglobin alpha c
879	33.5	32.2	183	2	A99262	probable acetyltra	952	33	31.7	141	1	HA1G1	hemoglobin alpha c
880	33.5	32.2	205	2	F36138	hypothetical prote	953	33	31.7	142	2	S04071	hemoglobin alpha c
881	33.5	32.2	205	2	A85604	urase accessory p	954	33	31.7	145	2	H89587	hemoglobin alpha c
882	33.5	32.2	205	2	G90794	probable urease ac	955	33	31.7	145	2	A41897	protein R09F10.5 [
883	33.5	32.2	205	2	S7462	urase accessory p	956	33	31.7	149	2	A26042	cellulase homolog
884	33.5	32.2	212	2	I46176	probable urease ac	957	33	31.7	149	2	A26042	globin I - sea lam
885	33.5	32.2	218	1	XURTG4	glutathione transf	958	33	31.7	153	2	E86872	galactoside O-acet
886	33.5	32.2	218	2	S01719	glutathione transf	959	33	31.7	157	2	G75560	conserved hypotnet
887	33.5	32.2	218	2	A23732	glutathione transf	960	33	31.7	157	2	B69817	hypothetical prote
888	33.5	32.2	218	2	B34159	glutathione transf	961	33	31.7	157	2	G69817	hypothetical prote
889	33.5	32.2	234	2	S32742	glutathione transf	962	33	31.7	158	1	A42082	protein-tyrosine-p
890	33.5	32.2	243	2	A36124	glutathione transf	963	33	31.7	159	2	T03958	heat shock protein
891	33.5	32.2	243	2	I54290	glutathione transf	964	33	31.7	160	2	E87384	transcription regu
892	33.5	32.2	255	1	HLHUD1	glutathione transf	965	33	31.7	162	2	A81325	probable signal-tr
893	33.5	32.2	266	2	A72401	avrFma1 protein -	966	33	31.7	164	2	B81325	conserved hypotnet
894	33.5	32.2	267	2	G70455	genome polyprotein	967	33	31.7	169	2	B72311	uncharacterized co
895	33.5	32.2	288	2	D87706	thyroglobulin 2 pr	968	33	31.7	171	2	A96901	secreted protein H
896	33.5	32.2	288	2	E90172	cell surface glyco	969	33	31.7	172	2	G82987	uncharacterized co
897	33.5	32.2	301	2	D81745	MHC class II histo	970	33	31.7	172	2	T06832	ycf37 protein - Cy
898	33.5	32.2	305	2	S38817	ABC transporter, A	971	33	31.7	172	2	T06832	hypothetical prote
899	33.5	32.2	320	2	A69789	hypothetical prote	972	33	31.7	176	2	S23343	rbbd protein - Shi
900	33.5	32.2	387	2	AI1050	diaminopelate ep	973	33	31.7	181	2	E55213	trpP-4-dehydroxam
901	33.5	32.2	411	2	T15705	hypothetical prote	974	33	31.7	183	2	S78545	hypothetical prote
902	33.5	32.2	416	2	D71936	hypothetical prote	975	33	31.7	183	2	E69432	hypothetical prote
903	33.5	32.2	417	2	JH0660	probable phosphati	976	33	31.7	184	2	AI3539	hypothetical prote
904	33.5	32.2	422	2	AG2999	FMRamide-related	977	33	31.7	185	2	B97443	hypothetical prote
905	33.5	32.2	422	2	B98284	fructokinase homol	978	33	31.7	185	2	A26661	conserved hypotnet
						979	33	31.7	189	2	D69389	DNA-directed RNA p	
						980	33	31.7	189	2	T17571	hypothetical prote	
						981	33	31.7	189	2	T17571	hypothetical prote	
						982	33	31.7	190	2	H64202	translation elonga	
						983	33	31.7	190	2	H64202	translation elonga	
						984	33	31.7	192	2	S36199	hypothetical prote	
						985	33	31.7	192	2	S36199	hypothetical prote	
						986	33	31.7	194	2	T44442	hypothetical prote	
						987	33	31.7	194	2	T44442	hypothetical prote	
						988	33	31.7	196	2	AE0077	conserved hypotnet	
						989	33	31.7	196	2	AE0077	conserved hypotnet	
						990	33	31.7	201	2	T37036	hypothetical prote	

979 33 31.7 202 2 A89911 conserved hypothet  
980 33 31.7 202 2 T28357 ORF MSV196 Ali mot  
981 33 31.7 202 2 F71293 hypothetical prote  
982 33 31.7 203 2 T41946 hypothetical prote  
983 33 31.7 219 2 AG2124 hypothetical prote  
984 33 31.7 220 2 AC0318 probable nicotinac  
985 33 31.7 223 2 F91210 hypothetical prote  
986 33 31.7 223 2 A86057 hypothetical prote  
987 33 31.7 224 2 D84345 phosphoribosylform  
988 33 31.7 230 1 WNWV26 p26 protein [simil  
989 33 31.7 233 2 A11152 hypothetical prote  
990 33 31.7 233 2 AB1512 hypothetical prote  
991 33 31.7 233 2 F82379 probable chenetaxi  
992 33 31.7 239 2 H90218 SSU ribosomal prot  
993 33 31.7 240 2 S71458 ABC-type heme tran  
994 33 31.7 240 2 S71456 ABC-type heme tran  
995 33 31.7 240 2 S62088 probable transport  
996 33 31.7 240 2 A43912 myogenin - Japanes  
997 33 31.7 242 2 A82754 lipoprotein XF0855  
998 33 31.7 243 2 A95162 alpha-acetolactate  
999 33 31.7 243 2 H98027 acetolactate decar  
1000 33 31.7 244 2 F84969 flagellar basal-bo

## ALIGNMENTS

RESULT 1  
JC7506 heparanase protein 2a - human  
C:Species: Homo sapiens (man)  
C>Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 09-Jul-2004  
C:Accession: JC7506  
R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hirccock, M.; Pa  
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000  
A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family m  
A:Reference number: JC7506  
A:Accession: JC7506  
A:Molecule type: mRNA  
A:Residues: 1-480 <MCK>  
A:Cross-references: UNIPROT:Q9HB39; UNIPARC:UPI0000003E88A; GB:AF282885  
C:Comment: This protein, a intracellular membrane-bound enzyme, has biological and thera  
therapies.  
C:Genetics:  
A:Gene: hpa2a  
A:Map position: 10q23-10q24  
C:Keywords: heparin binding; membrane bound

Query Match 63.5%; Score 66; DB 2; Length 480;  
Best Local Similarity 75.0%; Pred. No. 0.0074;  
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PAVLRFGGTTDFLIIF 16  
DB 110 PAVLRFGGKRTDFLQF 125  
||:|||||:|||||  
RESULT 2  
D70394 mannose6phosphatase A - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004  
C:Accession: D70394  
R:Beckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov  
V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: D70394  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-588 <AQF>  
A:Cross-references: UNIPROT:O67183; UNIPARC:UPI0000005651A; GB:AE000723; NID:g2983569; PJ

A:Experimental source: strain VFS  
C:Genetics:  
A:Gene: mtfa

Query Match 47.1%; Score 49; DB 2; Length 588;  
Best Local Similarity 60.0%; Pred. No. 5.5;  
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 YLRFGGTKTDFLIIF 17  
DB 307 YLRFKKTKTVFLVVD 321  
||:|||||:|||||

## RESULT 3

CRH07 carbonate dehydratase (EC 4.2.1.1) VII - human

N:Alternate names: carbonic anhydrase VII

C:Species: Homo sapiens (man)

C>Date: 10-Feb-1995 #sequence\_revision 05-May-1995 #text\_change 09-Jul-2004

C:Accession: A55272

R:Montgomery, J.C.; Venta, P.J.; Eddy, R.L.; Fukushima, Y.S.; Shows, T.B.; Tashian, R.E.

Genomics 11, 835-848, 1991

A:Title: Characterization of the human gene for a newly discovered carbonic anhydrase, C/

A:Reference number: A55272; MUID:92147127; PMID:1783392

A:Accession: A55272

A:Molecule type: DNA

A:Residues: 1-264 <MON>

A:Cross-references: UNIPROT:P43166; UNIPARC:UPI000000D814; GB:M76423; NID:gl79964; PIDN:/

A:Note: sequence extracted from NCBI backbone (NCBIN:80199, NCBIN:80201, NCBIN:80205, NCI

C:Genetics:

A:Gene: GDB:CA7

A:Cross-references: GDB:119741; OMIM:114770

A:Map position: 16q22.1-16q22.1

A:Introns: 14/1; 80/1; 119/3; 151/3; 172/3; 224/3

C:Function:

A:Description: catalyzes the reversible dissociation of carbonic acid to carbon dioxide

A:Note: this form is expressed in salivary gland and other tissues

C:Superfamily: carbonate dehydratase; carbonic anhydrase homology

C:Keywords: carbon-oxygen lyase; hydro-lyase; metalloprotein; zinc

F:2-264/Product: carbonate dehydratase VII #status predicted <MAT>

F:7-262/Domain: carbonic anhydrase homology <CAH>

F:96,98,121/Binding site: zinc (His) #status predicted

Query Match 46.2%; Score 48; DB 1; Length 264;

Best Local Similarity 56.2%; Pred. No. 3.5;

Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 LRFGGTTDFLIIFDPK 19  
DB 169 VRFGTKAQFSCFNPK 184  
||:|||||:|||||

## RESULT 4

C86516 hypothetical protein CPJ0203 [imported] - Chlamydomophila pneumoniae (strain J138)

C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004

C:Accession: C86516

R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise

Nucleic Acids Res. 28, 2311-2314, 2000

A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.

A:Reference number: A86491; MUID:20330349; PMID:10871362

A:Accession: C86516

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-265 <STO>

A:Cross-references: UNIPROT:Q928Y1; UNIPARC:UPI00000470E1; GB:BA000008; NID:g8978576; PII

A:Experimental source: strain J138

C:Genetics:

A:Gene: CPJ0203

Query Match 46.2%; Score 48; DB 2; Length 265;

Best Local Similarity 47.1%; Pred. No. 3.5;

Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 AYLRFGGKTDFLIFDP 18  
: ||||| :  
Db 24 SYFFFGGTRTQILVTP 40

RESULT 5  
KDO-transferase 2, probable frameshift CP0564 [imported] - Chlamydomydia pneumoniae (str  
C;Species: Chlamydomydia pneumoniae, Chlamydia pneumoniae  
C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C;Accession: D72105; D81563  
R;Kelman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A;Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A;Reference number: A72000; MUID:99206606; PMID:10192388  
A;Accession: D72105  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-265 <ARN>  
A;Cross-references: UNIPROT:Q9Z8Y1; UNIPARC:UPI00000470E1; GB:AE001363; NID  
A;Experimental source: strain CWL029  
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A;Reference number: A81500; MUID:20150255; PMID:10684935  
A;Accession: D81563  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-265 <REA>  
A;Cross-references: UNIPARC:UPI00000470E1; GB:AE002215; GB:AE002161; NID:g7189472; PIDN:  
A;Experimental source: strain AR39, HL cells  
C;Genetics:  
A;Gene: CPN0203; CP0564

Query Match 46.2%; Score 48; DB 2; Length 265;  
Best Local Similarity 47.1%; Pred. No. 3.5;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 AYLRFGGKTDFLIFDP 18  
: ||||| :  
Db 24 SYFFFGGTRTQILVTP 40

RESULT 6  
hypothetical protein TM0646 - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C;Accession: E72351  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.W.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.N.  
Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A;Reference number: A72200; MUID:99287316; PMID:10360571  
A;Accession: E72351  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-176 <ARN>  
A;Cross-references: UNIPROT:Q9WZB4; UNIPARC:UPI00000C13C4; GB:AE001738; GB:AE000512; NID  
A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TM0646  
C;Superfamily: Thermotoga maritima hypothetical protein TM0646

Query Match 45.2%; Score 47; DB 2; Length 176;  
Best Local Similarity 52.9%; Pred. No. 3.3;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 PAYLRFGGKTDFLIFD 17

Db 111 PKDARFIGHTPVDVFED 127  
||| ||| ||| :|||  
||| ||| ||| :|||

RESULT 7  
E90986  
probable colanic acid polymerase [imported] - Escherichia coli (strain O157:H7, substrain  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
gaawara, T.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
A;Reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: E90986  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-405 <HAY>  
A;Cross-references: UNIPROT:Q8X7M9; UNIPARC:UPI00000D0A2D; GB:BA000007; PIDN:BA836284.1;  
A;Experimental source: strain O157:H7, substrain RMD 050952  
C;Genetics:  
A;Gene: ECs2861

Query Match 44.2%; Score 46; DB 2; Length 405;  
Best Local Similarity 43.8%; Pred. No. 12;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 3 YLRFGGKTDFLIFDP 18  
: ||||| : |||  
Db 162 FIKFGGKRTALYFEP 177  
: ||||| : |||

RESULT 8  
H85831  
probable colanic acid polymerase wcad [imported] - Escherichia coli (strain O157:H7, sub  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C;Accession: H85831  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: H85831  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-405 <STO>  
A;Cross-references: UNIPROT:Q8X7M9; UNIPARC:UPI00000D0A2D; GB:AE005174; NID:g12516250; P  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: wcad

Query Match 44.2%; Score 46; DB 2; Length 405;  
Best Local Similarity 43.8%; Pred. No. 12;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 3 YLRFGGKTDFLIFDP 18  
: ||||| : |||  
Db 162 FIKFGGKRTALYFEP 177  
: ||||| : |||

RESULT 9  
G64971  
Putative colanic acid polymerase - Escherichia coli (strain K-12)  
C;Species: Escherichia coli  
C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 09-Jul-2004  
C;Accession: G64971  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: G64971  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-405 <BLAT>  
A;Cross-references: UNIPROT:P71238; UNIPARC:UPI0000138EC5; GB:AE000295; GB:U00096; NID:9  
A;Experimental source: strain K-12, substrain WGL655

Query Match 44.2%; Score 46; DB 2; Length 405;  
Best Local Similarity 43.8%; Pred. No. 12;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 3 YLRFGGTKTDFLIFDP 18  
::|||:|:|:  
Db 162 FIKFGKRTALYFEP 177

RESULT 10  
S77596  
cytochrome-c oxidase (EC 1.9.3.1) type cbb3 chain fixO - Paracoccus denitrificans  
C;Species: Paracoccus denitrificans  
C;Date: 24-Oct-1998 #sequence\_revision 24-Oct-1998 #text\_change 09-Jul-2004  
C;Accession: S77596  
R;de Gier, J.W.; Schepper, M.; Reijnders, W.N.M.; van Dyck, S.J.; Slotboom, D.J.; Warne,  
Mol. Microbiol. 20, 1247-1260, 1996  
A;Title: Structural and functional analysis of aa(3)-type and cbb(3)-type cytochrome c  
A;Reference number: S77595; MUID:96405647; PMID:8809776  
A;Accession: S77596  
A;Molecule type: DNA  
A;Residues: 1-241 <DEA>  
A;Cross-references: UNIPROT:Q51680; UNIPARC:UPI00000082F1; EMBL:U34353; NID:gl002874; PT  
A;Experimental source: strain Pd1222  
C;Genetics:  
A;Gene: ccoO  
C;Function:  
A;Pathway: respiratory chain  
C;Superfamily: Rhizobium cytochrome-c oxidase fixO chain  
C;Keywords: electron transfer; membrane-associated complex; oxidoreductase; respiratory

Query Match 43.3%; Score 45; DB 2; Length 241;  
Best Local Similarity 52.9%; Pred. No. 9.8;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 AYLRFGGTKTDFLIFDP 18  
|||:|:|:|:  
Db 221 AYLVLTGVTWDFSTFEP 237

RESULT 11  
B55582  
cytochrome-c oxidase (EC 1.9.3.1) fixO chain - Azorhizobium caulinodans  
N;Alternate names: cb-type cytochrome-c oxidase 28K chain; cytochrome b410; fixO protein  
C;Species: Azorhizobium caulinodans  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C;Accession: B55582; S42230  
R;Mandon, K.; Kaminski, P.A.; Elmerich, C.  
J. Bacteriol. 176, 2560-2568, 1994  
A;Title: Functional analysis of the fixNOQP region of Azorhizobium caulinodans.  
A;Reference number: A55582; MUID:94222833; PMID:8169204  
A;Accession: B55582  
A;Status: preliminary; nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 1-246 <MAN1>  
A;Cross-references: UNIPROT:Q43943; UNIPARC:UPI0000085F6A; GB:X74410; NID:9456310; PIDN:  
R;Mandon, K.; Kaminski, P.A.; Mougel, C.; Desnoues, N.; Dreyfus, B.; Elmerich, C.  
FEMS Microbiol. Lett. 114, 185-190, 1993  
A;Title: Role of the fixGHI region of Azorhizobium caulinodans in free-living and symbio  
A;Reference number: S42229; MUID:94109675; PMID:8282187  
A;Accession: S42230  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-246 <MAN2>  
A;Cross-references: UNIPARC:UPI0000085F6A; EMBL:X74410; NID:9456310; PIDN:CAA52430.1; PT  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1993

C;Superfamily: Rhizobium cytochrome-c oxidase fixO chain  
C;Keywords: electron transfer; membrane-associated complex; oxidoreductase; respiratory

Query Match 43.3%; Score 45; DB 1; Length 246;  
Best Local Similarity 50.0%; Pred. No. 10;  
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AYLRFGGKTDFLIFDPK 19  
Db 225 AYLQLGLGTQVDFKLYDNK 242  
|||: |||: |||: |  
|||: |||: |||: |

RESULT 12  
F84363  
Brp-like homolog [imported] - Halobacterium sp. NRC-1  
C;Species: Halobacterium sp. NRC-1  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C;Accession: F84363  
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithausner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li:  
A;Title: Genome sequence of Halobacterium species NRC-1.  
A;Reference number: A84160; MUID:20504483; PMID:11016950  
A;Accession: F84363  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-284 <STO>  
A;Cross-references: UNIPROT:Q9HNE6; UNIPARC:UPI0000063A63; GB:AE004437; NID:gl0581551; P1  
C;Genetics:  
A;Gene: blh

Query Match 42.3%; Score 44; DB 2; Length 284;  
Best Local Similarity 50.0%; Pred. No. 17;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 6 FGGTKTDFLIFDPK 19  
Db 118 FGGSVGLAVFDPR 131  
|||: |||: |||:  
|||: |||: |||:

RESULT 13  
D87403  
cbbC protein [imported] - Caulobacter crescentus  
C;Species: Caulobacter crescentus  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 05-Oct-2004  
C;Accession: D87403  
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonel, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A;Title: Complete Genome Sequence of Caulobacter crescentus.  
A;Reference number: A87249; MUID:21173698; PMID:11259647  
A;Accession: D87403  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-762 <STO>  
A;Cross-references: UNIPROT:Q9ASV8; UNIPARC:UPI000000C7309; GB:AE005673; NID:gl3422572; P1  
C;Genetics:  
A;Gene: CC1243  
C;Superfamily: molybdopterin dependent formate dehydrogenase

Query Match 42.3%; Score 44; DB 2; Length 762;  
Best Local Similarity 53.3%; Pred. No. 47;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 5 RFGGTKTDFLIFDPK 19  
Db 617 RTASGKANFLVFDPK 631  
|||: |||: |||:  
|||: |||: |||:

RESULT 14

JC6067  
CCAAT-binding factor CBF1 - mouse  
N:Alternate names: CBF1 protein  
C:Species: Mus musculus (house mouse)  
C>Date: 21-Jan-1997 #sequence\_revision 21-Jan-1997 #text\_change 09-Jul-2004  
R:Accession: JC6067; S65568; S65567  
R:Hoepfner, M.A.; Gilbert, D.J.; Copeland, N.G.; Jenkins, N.A.; Linzer, D.I.H.; Wu, B.  
Nucleic Acids Res. 24, 1091-1098, 1996  
A:Title: Cloning and characterization of mouse CCAAT binding factor.  
A:Reference number: S65567; MUID:96184859; PMID:8604343  
A:Accession: JC6067  
A:Molecule type: mRNA  
A:Residues: 1-997 <HOB>  
A:CROSS-references: UNIPROT:P53569; UNIPARC:UPI0000029995; GB:U19891; NID:G790574; PIDN:  
A:Accession: S65568  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-454, 'KR', 984-985, 'IQA' <HO2>  
A:CROSS-references: UNIPARC:UPI00000279A2; EMBL:U19892; NID:G790576; PIDN:AAB01504.1; PI  
C:Comment: This protein is important in mammalian cells for both growth promoting and gr  
C:Genetics:  
A:Gene: Cbf  
A:Map position: 17  
C:Keywords: alternative splicing; growth regulation  
F:943-967/Region: nuclear location signal

Query Match 42.3%; Score 44; DB 2; Length 997;  
Best Local Similarity 66.7%; Pred. No. 62;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 LRFGGTKTDFLI 15  
||| ||||| |||  
Db 46 LRLGGTKQDYLM 57

RESULT 15  
A36368  
transcription factor CBF, CCAAT-binding - human  
C:Species: Homo sapiens (man)  
C>Date: 28-Mar-1991 #sequence\_revision 28-Mar-1991 #text\_change 09-Jul-2004  
C:Accession: A36368  
R:Lum, L.S.Y.; Sultzman, L.A.; Kaufman, R.J.; Linzer, D.I.H.; Wu, B.J.  
Mol. Cell. Biol. 10, 6709-6717, 1990  
A:Title: A cloned human CCAAT-box-binding factor stimulates transcription from the human  
A:Reference number: A36368; MUID:91061780; PMID:2247079  
A:Accession: A36368  
A:Molecule type: mRNA  
A:Residues: 1-998 <LUM>  
A:CROSS-references: UNIPROT:Q03701; UNIPARC:UPI0000127180; GB:M37197; NID:G179968; PIDN:  
C:Genetics:  
A:Gene: GDB:CEBPA; CEBP  
A:CROSS-references: GDB:128839; OMIM:116897  
A:Map position: 19q13.1-19q13.1

Query Match 42.3%; Score 44; DB 2; Length 998;  
Best Local Similarity 66.7%; Pred. No. 62;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 LRFGGTKTDFLI 15  
||| ||||| |||  
Db 46 LRLGGTKQDYLM 57

RESULT 16  
S49346  
cytochrome-c oxidase (EC 1.9.3.1) fixO chain - Rhodobacter capsulatus  
N:Alternate names: cb-type cytochrome-c oxidase 28K chain; ccoO protein; cytochrome b410  
C:Species: Rhodobacter capsulatus  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: S65859; C54235; D54235; F54235; F54235; S49346  
R:Thoeny-Meyer, L.; Beck, C.; Preisig, O.; Hennecke, H.  
Mol. Microbiol. 14, 705-716, 1994  
A:Title: The ccoNOQP gene cluster codes for a cb-type cytochrome oxidase that functions

A:Reference number: S65858; MUID:95198544; PMID:7891558  
A:Accession: S65859  
A>Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-242 <THO>  
R:Gray, K.A.; Grooms, M.; Myllykallio, H.; Moomaw, C.; Slaughtcr, C.; Daldal, F.  
Biochemistry 33, 3120-3127, 1994  
A:Title: Rhodobacter capsulatus contains a novel cb-type cytochrome c oxidase without a c  
A:Reference number: A54235; MUID:94176508; PMID:8130227  
A:Accession: C54235  
A:Molecule type: protein  
A:Residues: 'XX', 3-6, 'XX', 9-20 <GRA1>  
A:CROSS-references: UNIPARC:UPI000017219C  
A:Experimental source: pMT0-404/MT-RBC1 cells  
A:Note: sequence extracted from NCBI backbone (NCBIP:144512)  
A:Accession: D54235  
A:Molecule type: protein  
A:Residues: 119, 'H', 121-128, 'S', 130-132, 'XSG', 136, 'F' <GRA2>  
A:CROSS-references: UNIPARC:UPI000017219D  
A:Experimental source: pMT0-404/MT-RBC1 cells  
A:Note: sequence extracted from NCBI backbone (NCBIP:144514)  
A:Accession: E54235  
A:Molecule type: protein  
A:Residues: 160-164, 'L', 166-170, 'YDAPFOAN' <GRA3>  
A:CROSS-references: UNIPARC:UPI000017219E  
A:Experimental source: pMT0-404/MT-RBC1 cells  
A:Note: sequence extracted from NCBI backbone (NCBIP:144517)  
A:Accession: F54235  
A:Molecule type: protein  
A:Residues: 180-192, 'A', 194, 'XAN' <GRA4>  
A:CROSS-references: UNIPARC:UPI000017219F  
A:Experimental source: pMT0-404/MT-RBC1 cells  
A:Note: sequence extracted from NCBI backbone (NCBIP:144519)  
C:Genetics:  
A:Gene: ccoO  
C:Function:  
A:Description: this cytochrome-c oxidase complex catalyzes the oxidation of four molecule  
A:Pathway: respiratory chain  
C:Superfamily: Rhizobium cytochrome-c oxidase fixO chain  
C:Keywords: electron transfer; membrane-associated complex; oxidoreductase; respiratory c

Query Match 41.3%; Score 43; DB 1; Length 242;  
Best Local Similarity 52.9%; Pred. No. 21;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 AYLRFGGKTDFLI 18  
|||: ||| |||  
Db 222 AYLVQVIGTWDFSTQP 238

RESULT 17  
B47468  
cytochrome-c oxidase (EC 1.9.3.1) fixO chain - Bradyrhizobium japonicum  
N:Alternate names: cb-type cytochrome-c oxidase 28K chain; cytochrome b410; fixO protein;  
C:Species: Bradyrhizobium japonicum  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: B47468  
R:Preisig, O.; Anthamatten, D.; Hennecke, H.  
Proc. Natl. Acad. Sci. U.S.A. 90, 3309-3313, 1993  
A:Title: Genes for a microaerobically induced oxidase complex in Bradyrhizobium japonicum  
A:Reference number: A47468; MUID:93234486; PMID:8386371  
A:Accession: B47468  
A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-244 <PRE>  
A:CROSS-references: UNIPROT:Q03074; UNIPARC:UPI0000011F85; GB:L07487; NID:G152196; PIDN:  
A:Experimental source: 1109pc4  
A:Note: sequence extracted from NCBI backbone (NCBIP:129654)  
C:Superfamily: Rhizobium cytochrome-c oxidase fixO chain  
C:Keywords: electron transfer; membrane-associated complex; oxidoreductase; respiratory c

Query Match 41.3%; Score 43; DB 1; Length 244;





C;Species: Methanobacterium thermoautotrophicum  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C;Accession: A69191  
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
i; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, M.;  
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func01  
A;Reference number: A69000; MUID:98037514; PMID:9371463  
A;Accession: A69191  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-240 <MTH>  
A;Cross-references: UNIPROT:O26779; UNIPARC:UPI0000129BAD; GB:AE000848; GB:AE000666; NID  
A;Experimental source: strain Delta H  
C;Genetics:  
A;Gene: MTH683  
A;Start codon: GTG  
C;Superfamily: tRNA nucleotidyltransferase  
C;Keywords: nucleotidyltransferase; tRNA processing

Query Match 40.4%; Score 42; DB 2; Length 240;  
Best Local Similarity 38.9%; Pred. No. 30;  
Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 2 AYLRFGGKTDFLIFDPK 19  
:|||||:|:  
Db 43 SYLEFGGNKILVAVYGR 60

RESULT 23  
H81327  
hypothetical protein Cj1214c [imported] - Campylobacter jejuni (strain NCTC 11168)  
C;Species: Campylobacter jejuni  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C;Accession: H81327  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barzel  
Nature 403, 665-668, 2000  
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A;Reference number: A81250; MUID:20150912; PMID:10688204  
A;Accession: H81327  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-241 <PAR>  
A;Cross-references: UNIPROT:O2PN81; UNIPARC:UPI000000CJ1E6C; GB:AL139077; GB:AL111168; NID  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: Cj1214c

Query Match 40.4%; Score 42; DB 2; Length 241;  
Best Local Similarity 47.1%; Pred. No. 30;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PAYLRFGGKTDFLIFD 17  
|||||:|:  
Db 204 PFYSRIGGYKSNFFDN 220

RESULT 24  
H72485  
probable hydantoinease APE2528 - Aeropyrum pernix (strain K1)  
C;Species: Aeropyrum pernix  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004  
C;Accession: H72485  
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999  
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy  
A;Reference number: A72450; MUID:99310339; PMID:10382966  
A;Accession: H72485  
A;Status: preliminary  
A;Molecule type: DNA

A;Residues: 1-587 <KAW>  
A;Cross-references: UNIPROT:Q9Y8V6; UNIPARC:UPI000005E3A4; DDBJ:AP000064; NID:gs5105945; PID:G  
A;Experimental source: strain K1  
C;Genetics:  
A;Gene: APE2528  
C;Superfamily: Pseudomonas D-amino acid hydantoin hydrolase (ATP-hydrolyzing) hyua

Query Match 40.4%; Score 42; DB 2; Length 687;  
Best Local Similarity 63.6%; Pred. No. 90;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 7 GGTKTDFLIED 17  
||| ||| :||  
Db 14 GGTFTDLIVED 24

RESULT 25  
A49672  
transcription factor Nrfl - human  
N;Alternate names: basic leucine-zipper transcription factor; NF-E2-related factor 1  
C;Species: Homo sapiens (man)  
C;Date: 07-Apr-1994 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C;Accession: A49672  
R;Chan, J.Y.; Han, X.L.; Kan, Y.W.  
Proc. Natl. Acad. Sci. U.S.A. 90, 11371-11375, 1993  
A;Title: Cloning of Nrfl, an NF-E2-related transcription factor, by genetic selection in  
A;Reference number: A49672; MUID:94068605; PMID:8248256  
A;Accession: A49672  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-742 <CHA>  
A;Cross-references: UNIPROT:Q14494; UNIPARC:UPI000002AF20  
A;Experimental source: erythroleukemia cell line K562  
A;Note: sequence extracted from NCBI backbone (NCBIN:140521, NCBIP:140522)  
C;Superfamily: human transcription factor TFC11; fos/jun DNA-binding domain homology  
F;618-659/Domain: fos/jun DNA-binding domain homology <FD>

Query Match 40.4%; Score 42; DB 2; Length 742;  
Best Local Similarity 53.8%; Pred. No. 97;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 7 GGTKTDFLIDPK 19  
|| | ||| :|  
Db 339 GGCSQDFLIFSPE 351

RESULT 26  
A55004  
transcription factor TFC11 - human  
N;Alternate names: LCR-F1 protein  
C;Species: Homo sapiens (man)  
C;Date: 11-Nov-1994 #sequence\_revision 11-Nov-1994 #text\_change 09-Jul-2004  
C;Accession: A55004; S48097  
R;Luna, L.; Joensuu, O.; Skartlien, A.H.; Pedetour, F.; Turc-Carel, C.; Prydz, H.; Kolst  
Genomics 22, 553-562, 1994  
A;Title: Molecular cloning of a putative novel human bZIP transcription factor on chromo  
A;Reference number: A55004; MUID:95095252; PMID:8001966  
A;Accession: A55004  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-772 <LUN>  
A;Cross-references: UNIPROT:Q14494; UNIPARC:UPI000012FFCF; GB:X77366; NID:gs41677; PID:G  
R;Caterina, J.J.; Donze, D.; Sun, C.W.; Ciavatta, D.J.; Townes, T.M.  
Nucleic Acids Res. 22, 2383-2391, 1994  
A;Title: Cloning and functional characterization of LCR-F1: a bZIP transcription factor t  
A;Reference number: S48097; MUID:94310069; PMID:8036168  
A;Accession: S48097  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 326-772 <CAT>  
A;Cross-references: UNIPARC:UPI000016A09C  
C;Genetics:  
A;Gene: GDB:TCF11

A;Cross-references: GDB:293921; OMIM:600115  
A;Map position: 17q22-17q22  
C;Superfamily: human transcription factor TFC11; fos/jun DNA-binding domain homology  
C;Keywords: DNA binding; leucine zipper; transcription factor  
F;648-689/Domain: fos/jun DNA-binding domain homology <FJD>  
Query Match 40.4%; Score 42; DB 2; Length 772;  
Best Local Similarity 53.8%; Pred. No. 1e+02;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 7 GGTGTDLFIFDPK 19  
||| ||| : |||  
Db 369 GGCQDFLLFSPE 381  
||| ||| : |||  
RESULT 27  
JC4169  
phosphoenolpyruvate carboxylase (EC 4.1.1.31) - Thermus sp.  
C;Species: Thermus sp.  
C;Date: 12-Oct-1995 #sequence\_revision 10-Nov-1995 #text\_change 20-Jun-2000  
C;Accession: JC4169; PC4043  
R;Nakamura, T.; Yoshioka, I.; Takahashi, M.; Toh, H.; Izui, K.  
J. Biochem. 118, 319-324, 1995  
A;Title: Cloning and sequence analysis of the gene for phosphoenolpyruvate carboxylase  
A;Reference number: JC4169; MUID:96064150; PMID:8543565  
A;Contents: No. 71  
A;Accession: JC4169  
A;Molecule type: DNA  
A;Residues: 1-857 <NAK>  
A;Cross-references: UNIPARC:UPI0000126F06; DDBJ:D42166; NID:g1061019; PIDN:BAA07723.1; E  
A;Accession: PC4043  
A;Molecule type: protein  
A;Residues: 26-43;120-145;198-212;214-232;270-292;295-319;417-444;801-825 <NA2>  
A;Cross-references: UNIPARC:UPI0000175FA6; UNIPARC:UPI0000175FA7; UNIPARC:UPI0000175FA8;  
FAD  
C;Comment: This enzyme catalyzes the reaction of phosphoenolpyruvate with HCO3- to form o  
erotic role by replenishing C4 dicarboxylic acids in the citric acid cycle. In C4 plants  
C;Genetics:  
A;Gene: ppc  
C;Superfamily: phosphoenolpyruvate carboxylase  
C;Keywords: carbon-carbon lyase; carboxy-lyase  
Query Match 40.4%; Score 42; DB 2; Length 857;  
Best Local Similarity 50.0%; Pred. No. 1.1e+02;  
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
QY 4 LRFGGTKTDLFIFDPK 19  
||| ||| : |||  
Db 390 LRLGGVHPDFLALSPE 405  
||| ||| : |||  
RESULT 28  
T24474  
hypothetical protein T04H1.3 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T24474  
R;Harris, B.  
submitted to the EMBL Data Library, August 1996  
A;Reference number: Z19896  
A;Accession: T24474  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-144 <WIL>  
A;Cross-references: UNIPROT:Q22176; UNIPARC:UPI000006113D; EMBL:D78200; PIDN:CAB01580.1;  
A;Experimental source: clone T04H1  
C;Genetics:  
A;Gene: CESP:T04H1.3  
A;Map position: 5  
A;Introns: 57/1  
C;Superfamily: Caenorhabditis hypothetical protein C40H1.5  
Query Match 39.4%; Score 41; DB 2; Length 144;

Best Local Similarity 66.7%; Pred. No. 26;  
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 8 GTKTDFLIFDPK 19  
||| ||| : |||  
Db 77 GTYDILTLDPK 88  
||| ||| : |||  
RESULT 29  
S76494  
hypothetical protein - Synechocystis sp. (strain PCC 6803)  
C;Species: Synechocystis sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C;Accession: S76494  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,  
DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
S.  
A;Reference number: S74322; MUID:97061201; PMID:8905231  
A;Accession: S76494  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-227 <KAN>  
A;Cross-references: UNIPROT:P74519; UNIPARC:UPI000012E6FD; EMBL:D90915; GB:AB001339; NID:  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C;Superfamily: Escherichia coli lipote-protein ligase lipB  
Query Match 39.4%; Score 41; DB 2; Length 227;  
Best Local Similarity 44.4%; Pred. No. 41;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;  
QY 1 PAYLRFGGTKTDLFIFDP 18  
||| ||| : |||  
Db 47 PVTLTGSGTKYLKFPD 64  
||| ||| : |||  
RESULT 30  
S76610  
hypothetical protein - Synechocystis sp. (strain PCC 6803)  
C;Species: Synechocystis sp.  
A;Variety: PCC 6803  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C;Accession: S76610  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,  
DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
S.  
A;Reference number: S74322; MUID:97061201; PMID:8905231  
A;Accession: S76610  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-262 <KAN>  
A;Cross-references: UNIPROT:Q55802; UNIPARC:UPI00000D3563; EMBL:D64004; GB:AB001339; NID:  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C;Superfamily: ompR protein; response regulator homology  
C;Keywords: phosphoprotein  
F;33-149/Domain: response regulator homology <RRH>  
F;88/Binding site: phosphate (Asp) (covalent) #status predicted  
Query Match 39.4%; Score 41; DB 1; Length 262;  
Best Local Similarity 56.2%; Pred. No. 48;  
Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 4 LRFGGTKTDLFIFDPK 19  
||| ||| : |||  
Db 103 LRFGGDNTFILLVS AK 118  
||| ||| : |||  
RESULT 31  
A72209

hypothetical protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1998 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: A72209  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing  
A:Reference number: A72209; PMID:99287316; PMID:10360571  
A:Accession: A72209  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-444 <ARN>  
A:Cross-references: UNIPROT:Q9X2D2; UNIPARC:UPI00000C11ED; GB:AE001818; GB:AE000512; NID:10360571  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM1812

Query Match 39.4%; Score 41; DB 2; Length 444;  
Best Local Similarity 66.7%; Pred. NO. 83;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 YLRFGGTKDPL 14  
||| ||| ||| |||  
Db 226 YLMSGGKSDFL 237

RESULT 32  
A86303  
hypothetical protein F17F16.4 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: A86303  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luross, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; PMID:21016719; PMID:11130712  
A:Accession: A86303  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-471 <STO>  
A:Cross-references: UNIPROT:Q9FWQ9; UNIPARC:UPI00000A3269; GB:AE005172; NID:g99554731; PI  
C:Genetics:  
A:Map position: 1

Query Match 39.4%; Score 41; DB 2; Length 471;  
Best Local Similarity 53.3%; Pred. NO. 88;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AYLRFGGKTDFLIF 16  
||| : ||| : ||  
Db 278 AYLAIGVPTDLKULF 292

RESULT 33  
A70362  
N-methylhydantoinase A - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004  
C:Accession: A70362  
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; PMID:98196666; PMID:9537320

Query Match 39.4%; Score 41; DB 2; Length 5069;  
Best Local Similarity 47.1%; Pred. NO. 1e+03;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

A:Accession: A70362  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-660 <AQF>  
A:Cross-references: UNIPROT:O66925; UNIPARC:UPI0000056417; GB:AE000703; NID:g2983287; PI  
A:Experimental source: strain VF5  
C:Genetics:  
A:Gene: hyuA  
C:Superfamily: Pseudomonas D-amino acid hydantoin hydrolase (ATP-hydrolyzing) hyuA

Query Match 39.4%; Score 41; DB 2; Length 660;  
Best Local Similarity 61.5%; Pred. NO. 1.3e+02;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 7 GGTKTDFLIFDPK 19  
||| ||| : ||  
Db 10 GGTFTDFVYWDGK 22

RESULT 34  
B72608  
probable hyuA APE1328 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004  
C:Accession: B72608  
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kawa DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum  
A:Reference number: A72450; PMID:99310339; PMID:10382966  
A:Accession: B72608  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-785 <KAW>  
A:Cross-references: UNIPROT:Q9YCC8; UNIPARC:UPI000005DEDD; DBJ:AP000061; NID:G5104821; I  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE1328  
C:Superfamily: Pseudomonas D-amino acid hydantoin hydrolase (ATP-hydrolyzing) hyuA

Query Match 39.4%; Score 41; DB 2; Length 785;  
Best Local Similarity 63.6%; Pred. NO. 1.5e+02;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 7 GGTKTDFLIFD 17  
||| ||| : ||  
Db 116 GGTFTDIVVFD 126

RESULT 35  
T17464  
rifamycin polyketide synthase modules 4-6 - Amycolatopsis mediterranei  
C:Species: Amycolatopsis mediterranei  
C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 09-Jul-2004  
C:Accession: T17464  
R:Schupp, T.  
submitted to the EMBL Data Library, December 1997  
A:Reference number: Z18802  
A:Accession: T17464  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-5069 <SCH>  
A:Cross-references: UNIPROT:O52789; UNIPARC:UPI0000055B02; EMBL:AJ223012; NID:e1227119; I  
A:Experimental source: strain LBG A3136  
C:Keywords: carrier protein  
F:1631-1702/Domain: acyl carrier protein homology <ACP1>  
F:3238-3309/Domain: acyl carrier protein homology <ACP2>  
F:4939-5010/Domain: acyl carrier protein homology <ACP3>

Query Match 39.4%; Score 41; DB 2; Length 5069;  
Best Local Similarity 47.1%; Pred. NO. 1e+03;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 AYLRFGGKTDFLIFDP 18  
||:||||:||||  
Db 1783 AYLRQGGFLHEALFDP 1799

## RESULT 36

B44827

FMRamide-like peptide - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 31-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004

C:Accession: B44827

R:Rosoff, M.L.; Burglin, T.R.; Li, C.

J. Neurosci. 12, 2356-2361, 1992

A:Title: Alternatively spliced transcripts of the flp-1 gene encode distinct FMRamide-1

A:Reference number: A44827; MUID:92300457; PMID:1607945

A:Accession: B44827

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-164 <ROS>

A:Cross-references: UNIPROT:P41855; UNIPARC:UPI000017B6C7

A:Note: sequence extracted from NCBI backbone (NCBIP:106594)

C:Keywords: neuropeptide

Query Match 38.9%; Score 40.5; DB 2; Length 164;  
Best Local Similarity 43.5%; Pred. No. 36;  
Matches 10; Conservative 3; Mismatches 5; Indels 5; Gaps 1;

QY 1 PAVLRFGGKTDFLIF-----DP 18  
|||||:||||:||||

Db 104 PNFILRFGSRDPNLFRLFGKAAADP 126

## RESULT 37

G98278

hypothetical protein AGR\_L\_2370 [imported] - Agrobacterium tumefaciens (strain C58, Cere

C:Species: Agrobacterium tumefaciens

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004

C:Accession: G98278

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,

A.: Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; MUID:21608551; PMID:11743194

A:Accession: G98278

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-99 <KUR>

A:Cross-references: UNIPROT:Q8U4W7; UNIPARC:UPI00000D2931; GB:AE007870; PIDN:AAK89753.1;

C:Genetics:

A:Gene: AGR\_L\_2370

A:Map position: linear chromosome

Query Match 38.5%; Score 40; DB 2; Length 99;  
Best Local Similarity 44.4%; Pred. No. 25;  
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 PAVLRFGGKTDFLIFDP 18  
|||||:||||:||||

Db 8 PAYLIQGGKLDITLAIP 25

## RESULT 38

T51959

hypothetical protein [imported] - Picea mariana

C:Species: Picea mariana

C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 09-Jul-2004

C:Accession: T51959

R:Perry, D.J.; Bousquet, J.

Genetics 149, 1089-1098, 1998

A:Title: Sequence-tagged-site (STS) markers of arbitrary genes. Development, characteriz

A:Reference number: Z25268; MUID:98278823; PMID:9611216

A:Accession: T51959

A:Status: preliminary; translated from GB/EMBL/DBEJ

A:Molecule type: mRNA

A:Residues: 1-141 <PER>

A:Cross-references: UNIPROT:O65048; UNIPARC:UPI00000A2B06; EMBL:AF051205; PIDN:AAAC32110.1

A:Experimental source: embryo

C:Genetics:

A:Note: Sb08

Query Match 38.5%; Score 40; DB 2; Length 141;  
Best Local Similarity 58.8%; Pred. No. 37;  
Matches 10; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 4 LRFG--GTTDFLIFDP 18  
|||||:||||:||||

Db 19 LRFGLAGVKSILTSHP 35

## RESULT 39

AD1156

hypothetical protein lmo0652 [imported] - Listeria monocytogenes (strain EGD-e)

C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004

C:Accession: AD1156

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,

.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Krest, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat

ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AD1156

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-178 <GLA>

A:Cross-references: UNIPROT:O8Y981; UNIPARC:UPI0000055170; GB:NC\_003210; PIDN:CAC98730.1;

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lmo0652

Query Match 38.5%; Score 40; DB 2; Length 178;  
Best Local Similarity 42.9%; Pred. No. 47;  
Matches 9; Conservative 3; Mismatches 5; Indels 4; Gaps 1;

QY 3 YLRFG---GTTDFLIFDPK 19  
|||||:||||:||||

Db 154 YLRGFGFLGTGNGFHLVQKK 174

## RESULT 40

AB2204

lipote-protein ligase B [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004

C:Accession: AB2204

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anal

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AB2204

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-221 <KUR>

A:Cross-references: UNIPROT:O8YSA4; UNIPARC:UPI000012E6E5; GB:BA000019; PIDN:BAE74884.1;

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr3185

C:Superfamily: Escherichia coli lipote-protein ligase lipB

Query Match 38.5%; Score 40; DB 2; Length 221;  
Best Local Similarity 44.4%; Pred. No. 59;  
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

```
Qy 1 PAYLRFGGTKTDFLIFDP 18
      ||| :|||
Db 53 PVTYLGQSSLEFLAKFDP 70

RESULT 41
AG3447
cytochrome-c oxidase (EC 1.9.3.1) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: AG3447
R:DelVecchio, V.G.; Kaputral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letese
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: ADJ252; PMID:11756688
A:Accession: AG3447
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-243 <KUR>
A:Cross-references: UNIPROT:Q8YFF7; UNIPROT:Q8G2G0; UNIPARC:UPI0000058103; GB:AE008917;
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI1565
A:Map position: 1
C:Superfamily: Rhizobium cytochrome-c oxidase fixO chain
C:Keywords: oxidoreductase

Query Match 38.5%; Score 40; DB 2; Length 243;
Best Local Similarity 50.0%; Pred. No. 65;
Matches 10; Conservative 2; Mismatches 6; Indels 2; Gaps 1;

Qy 2 AYLRFGGKTDFLIFD--PK 19
      ||| :|||
Db 222 AYLQMLGLTVDFSTYDQSPK 241

RESULT 42
H97341
dihydrodipicolinate synthase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: H97341
R:Noelling, J.; Berton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; PMID:21359325; PMID:21359325
A:Accession: H97341
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-286 <KUR>
A:Cross-references: UNIPROT:Q97D80; UNIPARC:UPI0000128E35; GB:AE001437; PIDN:AAK81523.1;
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3600
C:Superfamily: dihydrodipicolinate synthase

Query Match 38.5%; Score 40; DB 2; Length 286;
Best Local Similarity 42.1%; Pred. No. 77;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFLIFDPK 19
      ||| :|||
Db 72 PIYVFGGNNTKMTKD1K 90

RESULT 43
T35417
probable beta-lactamase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004

C:Accession: T35417
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z21577
A:Accession: T35417
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-310 <OLI>
A:Cross-references: UNIPROT:Q9X7X2; UNIPARC:UPI000000DAF52; EMBL:AL049485; PIDN:CAB39710.1
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC6A5.26c

Query Match 38.5%; Score 40; DB 2; Length 310;
Best Local Similarity 57.1%; Pred. No. 83;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFL 14
      ||| :|||
Db 86 PQIPQFAGLKTDFL 99

RESULT 44
D89102
protein F25E5.2 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: D89102
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; PMID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C.eleg
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: D89102
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-338 <STO>
A:Cross-references: UNIPROT:O76655; UNIPARC:UPI000007D193; GB:chr_V; PIDN:AAC27333.1;
C:Genetics:
A:Gene: F25E5.2
A:Map position: 5

Query Match 38.5%; Score 40; DB 2; Length 338;
Best Local Similarity 50.0%; Pred. No. 91;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 3 YLRFGGTKTDFLIF 16
      ||| :|||
Db 257 FLKFGGSKEDLAKF 270

RESULT 45
AC0382
probable Pyridoxal-phosphate dependent protein YP03147 [imported] - Yersinia pestis (str
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C:Accession: AC0382
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; PMID:21470413; PMID:11586360
A:Accession: AC0382
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-346 <KUR>
A:Cross-references: UNIPROT:Q8ZC75; UNIPARC:UPI000000DC7C3; GB:AL590842; PIDN:CAC92382.1;
C:Genetics:
A:Gene: YP03147

Query Match 38.5%; Score 40; DB 2; Length 346;
```

RESULT 48  
S56954  
protein YBR162c homolog YJL171c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: probable membrane protein YJL171c; protein J0512  
C:Species: Saccharomyces cerevisiae  
C>Date: 05-May-1995 #sequence\_revision 08-Sep-1995 #text\_change 09-Jul-2004  
C:Accession: S56954  
R:Obermaier, B.; Piravandi, E.; Rinke, M.; Domdey, H.  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56937  
A:Accession: S56954  
A:Molecule type: DNA  
A:Residues: 1-396 <OBE>  
A:Cross-references: UNIPROT:P46992; UNIPARC:UPI000013B5FD; EMBL:Z49446; NID:910  
C:Genetics:  
A:Gene: MIPS:YJL171c

A:Map Position: 10L  
C:Keywords: transmembrane protein

Query Match	38.5%	Score 40;	DB 2;	Length 396;
Best Local Similarity	57.1%;	Pred. No. 1.le+02;		
Matches 8;	Conservative	1;	Mismatches 5;	Indels 0;
Gaps	0;			

A:Accession: 120314  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-419 <MWL>  
A:Cross-references: UNIPROT:O01320; UNIPARC:UPT000007F05F; EMBL:Z81494; P1DN:CA  
A:Experimental source: Clone F02E9  
C:Genetics:  
A:Gene: CESP:F02B9.7  
A:Map position: 1  
A:Introns: 25/3; 56/1; 90/1; 140/3; 257/2; 306/3; 347/2

RESULT 50  
G71657  
folypolyglutamate synthase (folC) RP536 - Rickettsia prowazekii  
C:Species: Rickettsia prowazekii  
C:Date: 21 Nov 1998 #sequence revision 21-Nov-1998 #text change 09-Jul-2004



C;Accession: G71657  
R;Anderson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichert-Ponten, T.; Alsmark, U.  
Nature 396, 133-140, 1998  
A;Title: The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria.  
A;Reference number: A71630; MUID:99039499; PMID:9823893  
A;Accession: G71657  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-433 <AND>  
A;Cross-references: UNIPROT:Q9ZD14; UNIPARC:UPI00000D37C8; GB:AJ235272; GB:AJ235269; NID  
A;Experimental source: strain Madrid E  
C;Genetics:  
A;Gene: folC; RP536  
C;Superfamily: folylpolyglutamate synthase

Query Match 38.5%; Score 40; DB 2; Length 433;  
Best Local Similarity 50.0%; Pred. No. 1.2e+02;  
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy	2	AYLRFGG	KTDFL	FD	17
			:		:
Db	121	AFLAFAQT	KADIL	ILE	136

Search completed: June 5, 2006, 12:53:28  
Job time : 39.0822 secs

OM protein - protein search, using sw model  
 Run on: June .5, 2006, 12:32:17 ; Search time 166.575 Seconds  
 (without alignments)  
 105.510 Million cell updates/sec

Title: US-10-645-659A-7  
 Perfect score: 104  
 Sequence: 1 PAYLRFQGTDFLIFDPK 19

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 1000 summaries

Database : UniProt 7.2.\*  
 1: uniprot\_sprot.\*  
 2: uniprot\_crembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	100.0	543	1 HPSE_HUMAN	Q9Y251 homo sapien
2	104	100.0	558	2 Q33X5 SPAJD	Q33X5 spalax juda
3	104	100.0	574	2 Q33X6 SPAJD	Q33X6 spalax juda
4	104	100.0	574	2 Q33X7 GRODE	Q33X7 spalax carm
5	104	100.0	574	2 Q33X8 GRODE	Q33X8 spalax gela
6	104	100.0	574	2 Q33X9 GRODE	Q33X9 spalax gali
7	99	95.2	535	1 HPSE_MOUSE	Q6Y9Z1 mus musculus
8	99	95.2	535	1 HPSE_RAT	Q71RP1 rattus norv
9	92	88.5	545	1 HPSE_BOVIN	Q9MYV0 bos taurus
10	81	77.9	523	1 HPSE_CHICK	Q9DYK5 gallus gall
11	75	72.1	533	2 Q4SYF6 TTENG	Q4SYF6 tetraodon n
12	66	63.5	592	1 HPSE2_HUMAN	Q8WWQ2 homo sapien
13	66	63.5	592	2 Q2M1H9 HUMAN	Q2M1H9 homo sapien
14	66	63.5	597	2 Q4TB80 TTENG	Q4TB80 tetraodon n
15	53	51.0	326	2 Q7MH05 VIBVY	Q7MH05 vibrio vuln
16	53	51.0	326	2 Q8DCU6 VIBVU	Q8DCU6 vibrio vuln
17	53	51.0	1538	2 Q2Q6E6 LEPIC	Q2Q6E6 leptospira
18	53	51.0	1538	2 Q8F5E5 LEPIN	Q8F5E5 leptospira
19	50	48.1	512	2 Q8LHX2 ORYSA	Q8LHX2 oryza sativ
20	49.5	47.6	224	2 Q4HLV4 CAMLA	Q4HLV4 campylobact
21	49	47.1	513	2 Q6Z4S8 ORYSA	Q6Z4S8 oryza sativ
22	49	47.1	588	2 Q67183 AQUAE	Q67183 aquifex aeo
23	48	46.2	161	2 Q8PS30 METMA	Q8PS30 methanosarc
24	48	46.2	176	2 Q8TKK5 METAC	Q8TKK5 methanosarc
25	48	46.2	194	2 Q2J1Z0 9CYAN	Q2J1Z0 cyanobacter
26	48	46.2	208	2 Q85YU0 HUMAN	Q85YU0 homo sapien
27	48	46.2	247	2 Q2RP20 RHORU	Q2RP20 rhodospiril
28	48	46.2	264	1 CAH7_HUMAN	P43166 homo sapien
29	48	46.2	265	2 Q9Z8Y1 CHLPN	Q9Z8Y1 chlamydia p
30	48	46.2	311	2 Q41BD7 9BACI	Q41BD7 exigubacte
31	48	46.2	411	2 Q8G431_BIFLO	Q8G431 bifidobacte

32	48	46.2	425	2	Q3FBC5_9BURK	Q3FBC5 burkholderi
33	48	46.2	515	2	Q8T108_BOMMO	Q8T108 bombyx mori
34	48	46.2	638	2	Q47YD5_COLP3	Q47YD5 colwellia p
35	48	46.2	742	2	Q8KZS8_ACEPA	Q8KZS8 acetobacter
36	48	46.2	838	2	Q2IRU0_RHOPA	Q2IRU0 rhodopseudo
37	48	46.2	1895	2	Q3QLR3_9RHOB	Q3QLR3 silicibacte
38	48	46.2	2405	2	Q2LZG0_DROPS	Q2LZG0 drosophila
39	47	45.2	176	2	Q9WZB4_THEMA	Q9WZB4 thermotoga
40	47	45.2	1076	2	Q5ZIQ7_CHICK	Q5ZIQ7 gallus gall
41	47	45.2	1076	2	Q5ZMV4_GALL	Q5ZMV4 gallus gall
42	46	44.2	180	2	Q33FF2_METHU	Q33FF2 methanospir
43	46	44.2	219	2	Q4CAE2_CROWT	Q4CAE2 crocospheer
44	46	44.2	308	2	Q5TMQ4_ANOGA	Q5TMQ4 anopheles g
45	46	44.2	319	2	Q2W843_MAGSA	Q2W843 magnetospir
46	46	44.2	319	2	Q2WAD9_MAGSA	Q2WAD9 magnetospir
47	46	44.2	404	2	Q57MP2_SALCH	Q57MP2 salmonella
48	46	44.2	404	2	Q5PDY6_SALPA	Q5PDY6 salmonella
49	46	44.2	404	2	Q9F7A7_SALTY	Q9F7A7 salmonella
50	46	44.2	405	1	WCAD_ECOLI	P71238 escherichia
51	46	44.2	405	2	Q323F6_SHIBS	Q323F6 shigella bo
52	46	44.2	405	2	Q32OE7_SHISS	Q32OE7 shigella so
53	46	44.2	405	2	Q2MAY3_ECOLI	Q2MAY3 escherichia
54	46	44.2	405	2	Q8X7M9_ECO57	Q8X7M9 escherichia
55	46	44.2	414	2	Q7XIE4_ORYSA	Q7XIE4 oryza sativ
56	46	44.2	429	2	Q8FG19_ECOL6	Q8FG19 escherichia
57	46	44.2	445	2	Q9ANW0_ECO57	Q9ANW0 escherichia
58	46	44.2	507	2	Q8LHX1_ORYSA	Q8LHX1 oryza sativ
59	46	44.2	527	2	Q9LRC8_SCUBA	Q9LRC8 scutellaria
60	46	44.2	536	2	Q2UDS9_ASPOR	Q2UDS9 aspergillus
61	46	44.2	1245	2	Q9YGH8_SCOMX	Q9YGH8 scophthalmu
62	45.5	43.8	251	1	SFSA_SYMTH	Q67GV9 symbiobacte
63	45.5	43.8	385	2	Q5S2S8_DICDI	Q5S2S8 dictyosteli
64	45.5	43.8	385	2	Q861C3_DICDI	Q861C3 dictyosteli
65	45.5	43.8	486	2	Q3DX84_CHLAU	Q3DX84 chloroflexu
66	45	43.3	183	2	Q64Z07_BACFR	Q64Z07 bacteroides
67	45	43.3	212	2	Q8BIS4_MOUSE	Q8BIS4 mus musculu
68	45	43.3	241	2	Q3PHF0_PARDE	Q3PHF0 paracoccus
69	45	43.3	241	2	Q51680_PARDE	Q51680 paracoccus
70	45	43.3	246	2	Q43943_AZOCA	Q43943 azorhizobiu
71	45	43.3	250	2	Q942D6_ORYSA	Q942D6 oryza sativ
72	45	43.3	451	2	Q8ESC3_OCBIH	Q8ESC3 oceanobacil
73	45	43.3	451	2	Q691I6_ORYSA	Q691I6 oryza sativ
74	45	43.3	557	2	Q4J5V6_AZOVI	Q4J5V6 azotobacter
75	45	43.3	591	2	Q2W0LI_MAGSA	Q2W0LI magnetospir
76	45	43.3	1006	2	Q5U4R2_XENLA	Q5U4R2 xenopus lae
77	45	43.3	1096	2	Q616F9_CAEBR	Q616F9 caenorhabdi
78	45	43.3	1623	2	Q8K449_MOUSE	Q8K449 mus musculu
79	45	43.3	3842	2	Q7R0N8_GIALA	Q7R0N8 giardia lam
80	44.5	42.8	606	2	Q2N8C9_9SPHN	Q2N8C9 erythroba
81	44.5	42.8	750	2	Q3RSB0_RALME	Q3RSB0 ralstonia m
82	44	42.3	136	2	Q3UUZ3_MOUSE	Q3UUZ3 mus musculu
83	44	42.3	190	2	Q979H0_THEVO	Q979H0 thermoplasm
84	44	42.3	284	2	Q9HNE6_HALSA	Q9HNE6 halobacteri
85	44	42.3	285	2	Q46JDI_PROMT	Q46JDI prochloroco
86	44	42.3	303	2	Q81QF9_BACAN	Q81QF9 bacillus an
87	44	42.3	305	2	Q6F2A4_MESFL	Q6F2A4 mesoplasma
88	44	42.3	435	2	Q8K2R7_MOUSE	Q8K2R7 mus musculu
89	44	42.3	536	2	Q2UT15_ASPOR	Q2UT15 aspergillus
90	44	42.3	568	2	Q5ZKA0_CHICK	Q5ZKA0 gallus gall
91	44	42.3	606	2	Q8LHM4_ORYSA	Q8LHM4 oryza sativ
92	44	42.3	652	2	Q4E3C0_TRYCR	Q4E3C0 trypanosoma
93	44	42.3	762	2	Q9A8V8_CAOCR	Q9A8V8 caulobacter
94	44	42.3	772	2	Q5RA25_PONPY	Q5RA25 pongo pygma
95	44	42.3	961	2	Q3UWQ7_MOUSE	Q3UWQ7 mus musculu
96	44	42.3	997	1	CEBPZ_MOUSE	CEBPZ mus musculu
97	44	42.3	998	1	CEBPZ_HUMAN	CEBPZ homo sapien
98	44	42.3	1002	2	Q2JQ83_9CYAN	Q2JQ83 cyanobacter
99	44	42.3	1040	2	Q3TXW5_MOUSE	Q3TXW5 mus musculu
100	44	42.3	1052	2	Q3TYA8_MOUSE	Q3TYA8 mus musculu
101	44	42.3	1054	2	Q8NE75_HUMAN	Q8NE75 homo sapien
102	44	42.3	1060	2	Q7SFT5_NEUCR	Q7SFT5 neurospora
103	44	42.3	1369	2	Q8UW86_PAROL	Q8UW86 paralichthy
104	44	42.3	2049	2	Q5W6T2_ORYSA	Q5W6T2 oryza sativ

105	43.5	41.8	136	2	Q4FM10_PELUB	Q4fm10 pelagibacte	178	42	40.4	345	2	Q59AE5_PRUPE	Q59ae5 prunus pers
106	43.5	41.8	318	2	Q7UFM0_RHOBA	Q7ufm0 rhodopirell	179	42	40.4	358	2	Q4JC72_SULAC	Q4jc72 sulfobolus
107	43.5	41.8	380	2	Q6CBE6_YARLI	Q6cbe6 yarrowia li	180	42	40.4	374	2	Q9P9G0_METMA	Q9p9g0 methanosear
108	43.5	41.8	432	2	Q2RPF9_RHORU	Q2rpf9 rhodospiril	181	42	40.4	394	2	Q7Q6S7_ANOGA	Q7q6s7 anopheles g
109	43.5	41.8	777	2	Q46Q10_RALEU	Q46q10 ralstonia e	182	42	40.4	402	2	Q8CLZ6_OCEIH	Q8clz6 oceanobacil
110	43	41.3	119	2	Q59Z68_CANAL	Q59z68 candida alb	183	42	40.4	402	2	Q8EN48_OCEIH	Q8en48 oceanobacil
111	43	41.3	199	2	Q9BI00_GLOPA	Q9bi00 globodera p	184	42	40.4	402	2	Q8EQP8_OCEIH	Q8eqp8 oceanobacil
112	43	41.3	238	2	Q4HQ48_CAMUP	Q4hq48 campylobact	185	42	40.4	428	1	ILL3_ARATH	ILL3 arabadopsis
113	43	41.3	242	2	Q30728_RHOCA	Q30728 rhodobacter	186	42	40.4	430	2	Q4HJR7_CAMLA	Q4hj7 campylobact
114	43	41.3	242	2	Q52687_RHOCA	Q52687 rhodobacter	187	42	40.4	432	2	Q5B5T2_EMENI	Q5b5t2 aspergillus
115	43	41.3	244	2	Q35QC5_BRAD	Q35qc5 bradyrhizob	188	42	40.4	445	1	HLX3_NOCFA	Q5v7q0 nocardia fa
116	43	41.3	244	2	Q30374_BRAJA	Q30374 bradyrhizob	189	42	40.4	467	2	Q6CGN1_YARLI	Q6cgn1 yarrowia li
117	43	41.3	244	2	Q6NDT6_RHOPA	Q6ndt6 rhodopseude	190	42	40.4	473	2	Q8G0U7_ORYSA	Q8g0u7 oryza sativ
118	43	41.3	252	1	SOJ_TREPA	Q83296 treponema p	191	42	40.4	505	2	Q3CZ02_STRAG	Q3cz02 streptococc
119	43	41.3	262	2	Q3Y546_PETMA	Q3y546 petromyzon	192	42	40.4	505	2	Q3D7K4_STRAG	Q3d7k4 streptococc
120	43	41.3	276	2	Q6DGN8_BRARE	Q6dgn8 brachydanio	193	42	40.4	505	2	Q3DDC2_STRAG	Q3ddc2 streptococc
121	43	41.3	294	2	Q5E2D1_VIBF1	Q5e2d1 vibrio fisc	194	42	40.4	505	2	Q3DL57_STRAG	Q3dl57 streptococc
122	43	41.3	364	2	Q38CI5_9TRYF	Q38ci5 trypanosoma	195	42	40.4	505	2	Q3JZ83_STRAI	Q3jz83 streptococc
123	43	41.3	452	2	Q4ICE6_GIBZE	Q4ice6 gibberella	196	42	40.4	505	2	Q8DXP7_STRAS	Q8dxdp7 streptococc
124	43	41.3	477	2	Q2QMV6_ORYSA	Q2qmv6 oryza sativ	197	42	40.4	514	2	Q8E3B7_STRAS	Q8e3b7 streptococc
125	43	41.3	502	1	XYLB_LACBR	Q35850 lactobacill	198	42	40.4	514	2	Q55FH8_PCEPI	Q55fh8 dictyosteli
126	43	41.3	552	2	Q4N4Z2_THEPA	Q4n4w2 theileria p	199	42	40.4	565	2	Q3KEN8_PSEPF	Q3ken8 pseudomonas
127	43	41.3	592	2	Q75F84_ASHGO	Q75f84 ashbya goss	200	42	40.4	620	2	Q3E362_CHLAU	Q3e362 chloroflexu
128	43	41.3	682	2	Q37Z83_NOVOPH	Q37z83 novosphingo	201	42	40.4	675	2	Q3IW73_RHOS4	Q3iw73 rhodobacter
129	43	41.3	796	2	Q33RG4_9CAMM	Q33rg4 shewanella	202	42	40.4	687	2	Q9Y8V6_AERPE	Q9y8v6 aeropyrum p
130	43	41.3	805	2	Q3GJ58_CHLVI	Q3gj58 prosthecoc	203	42	40.4	714	2	Q89P62_BRAJA	Q89p62 bradyrhizob
131	43	41.3	909	2	Q51B12_BNTHI	Q51b12 entamoeba h	204	42	40.4	743	2	Q3CTA8_ALTAT	Q3cta8 pseudolater
132	43	41.3	924	2	Q3AGS6_SYNSC	Q3ags6 synecococc	205	42	40.4	760	2	Q7RMA1_PLAYO	Q7rma1 plasmodium
133	43	41.3	1293	2	Q42511_ARATH	Q4251 arabidopsis	206	42	40.4	764	2	Q9M8J3_ARATH	Q9m8j3 arabidopsis
134	43	41.3	1327	2	Q5VKQ3_SACER	Q5vkq3 saccharopol	207	42	40.4	772	1	NFLJ1_HUMAN	Q14494 h nuclear f
135	42.5	40.9	164	2	Q23805_CAEVU	Q23805 caenorhabdi	208	42	40.4	791	2	Q8NF22_HUMAN	Q8nf22 homo sapien
136	42.5	40.9	164	2	Q7JKL0_CAEVL	Q7jkl0 caenorhabdi	209	42	40.4	809	2	Q9P9C4_9EURY	Q9p9c4 uncultured
137	42.5	40.9	174	2	Q23806_CAEVU	Q23806 caenorhabdi	210	42	40.4	830	2	Q6SHK6_9BACT	Q6shk6 uncultured
138	42.5	40.9	175	1	FARP_CAEVL	P41855 caenorhabdi	211	42	40.4	857	1	CAPP_THES7	P51060 thermus sp.
139	42.5	40.9	176	2	Q6QKQ3_CAEBR	Q6qkq3 caenorhabdi	212	42	40.4	917	2	Q6FM76_CANGA	Q6fm76 candida gla
140	42.5	40.9	249	2	Q74312_MYCPA	Q74312 mycobacteri	213	42	40.4	922	2	Q4N844_CHLLI	Q4n844 chlorobium
141	42.5	40.9	599	2	Q7X669_ORYSA	Q7x669 oryza sativ	214	42	40.4	1028	2	Q4WFC6_ASPFU	Q4wfc6 aspergillus
142	42.5	40.9	770	2	Q9LDM8_SOYBN	Q9ldm8 glycine max	215	42	40.4	1056	2	Q36MU3_MARHY	Q36mu3 marinobacte
143	42	40.4	90	2	Q6VW77_ASCSU	Q6v7w7 ascaris suu	216	42	40.4	1177	2	Q7S179_ORYSA	Q7s179 oryza sativ
144	42	40.4	108	2	Q2TFP0_WELIC	Q2tfp0 meloidogyne	217	42	40.4	1182	2	Q3GMT2_9GAMM	Q3gmt2 psychrobact
145	42	40.4	120	2	Q6TLU2_ASCSU	Q6tluj2 ascaris suu	218	42	40.4	1486	2	Q3XPD6_ORYSA	Q3xpd6 oryza sativ
146	42	40.4	123	2	Q58LR1_9CAUD	Q58lr1 cyanophag	219	42	40.4	1510	2	Q7XLY1_ORYSA	Q7xly1 oryza sativ
147	42	40.4	134	2	Q86Q05_9BILA	Q86qg5 oesophagost	220	42	40.4	1533	2	Q2R3J9_ORYSA	Q2r3j9 oryza sativ
148	42	40.4	139	2	Q9BH15_GLOPA	Q9bh15 globodera p	221	42	40.4	1534	2	Q4RQY9_TETNG	Q4rqy9 tetraodon n
149	42	40.4	143	2	Q9XWV7_CAEVL	Q9xwv7 caenorhabdi	222	42	40.4	1639	2	Q2QSC3_ORYSA	Q2qsc3 oryza sativ
150	42	40.4	149	2	Q86G60_HETGL	Q86g60 heterodera	223	42	40.4	1718	2	Q53Q06_ORYSA	Q53q06 oryza sativ
151	42	40.4	156	2	Q9QKCI_CANAL	Q9qkci candida alb	224	42	40.4	1860	2	Q2RB82_ORYSA	Q2rb82 oryza sativ
152	42	40.4	156	2	Q615A5_CAEBR	Q615a5 caenorhabdi	225	42	40.4	1890	2	Q2R1D8_ORYSA	Q2qr1d8 oryza sativ
153	42	40.4	172	1	VP19_PENSV	Q709n3 pelargonium	226	42	40.4	1924	2	Q2QRV1_ORYSA	Q2qrv1 oryza sativ
154	42	40.4	181	2	Q3X0Z7_9ACTN	Q3x0z7 rubrobacter	227	42	40.4	2266	2	Q3LJ77_ORYSA	Q3lj77 oryza sativ
155	42	40.4	181	2	Q4AP30_9CHLB	Q4ap30 chlorobium	228	42	40.4	2987	2	Q7QWR0_GIALA	Q7qwr0 giardia lam
156	42	40.4	207	2	Q5MF94_9WAGN	Q5mf94 piper betle	229	41.5	39.9	151	2	Q86Q04_9BILA	Q86q04 oesophagost
157	42	40.4	210	1	KGUA_BLOFL	Q7vrj7 blochmannia	230	41.5	39.9	177	2	Q3EZP8_BACTI	Q3ezp8 bacillus th
158	42	40.4	221	2	Q3HEC6_TRIER	Q3hec6 trichodesmi	231	41.5	39.9	184	2	Q8WUX2_HUMAN	Q8wux2 homo sapien
159	42	40.4	230	2	Q4HH06_CAMCO	Q4hh06 campylobact	232	41.5	39.9	184	2	Q8WV18_HUMAN	Q8wv18 homo sapien
160	42	40.4	236	2	Q5HTQ0_CAMJR	Q5htq0 campylobact	233	41.5	39.9	184	2	Q5PPV4_XENLA	Q5ppv4 xenopus lae
161	42	40.4	240	1	EXX1_METH	Q26779 methanobact	234	41.5	39.9	302	2	Q2T1U3_BURTH	Q2t1u3 burkholderi
162	42	40.4	240	2	Q40HU5_9RHOB	Q40hu5 jannaschia	235	41.5	39.9	327	2	Q4AK87_9CHLB	Q4ak87 chlorobium
163	42	40.4	241	2	Q9PN81_CAMJE	Q9pn81 campylobact	236	41.5	39.9	485	2	Q45771_BACTN	Q45771 bacteroides
164	42	40.4	242	2	Q3QB1_9RHOB	Q3qgb1 silicibacte	237	41.5	39.9	492	2	Q4V812_XENLA	Q4v812 xenopus lae
165	42	40.4	242	2	Q5LMN8_SILPO	Q5lmn8 silicibacte	238	41.5	39.9	509	2	Q6FMF2_CANGA	Q6fmf2 candida gla
166	42	40.4	251	2	Q73Q37_TREDE	Q73q37 treponema d	239	41.5	39.9	555	2	Q641A9_XENLA	Q641a9 xenopus lae
167	42	40.4	253	2	Q8ES45_OCEIH	Q8es45 oceanobacil	240	41.5	39.9	555	2	Q7SDV0_NEUCR	Q7sdv0 neurospora
168	42	40.4	273	2	Q9JRQ8_ACTAC	Q9jrq8 actinobacil	241	41.5	39.9	743	2	Q7U3G8_SYNPX	Q7u3g8 synecococc
169	42	40.4	280	2	Q3EHT3_ACTSC	Q3eht3 actinobacil	242	41.5	39.9	763	2	Q6EKW1_XENLA	Q6ekw1 xenopus lae
170	42	40.4	281	2	Q3DTE0_STRAG	Q3dte0 streptococc	243	41.5	39.9	763	2	Q6XV80_XENLA	Q6xv80 xenopus lae
171	42	40.4	294	2	Q6D207_ERWCT	Q6d207 erwinia car	244	41.5	39.9	793	2	Q5V302_HALMA	Q5v302 haloarcula
172	42	40.4	299	2	Q52QR4_SOYEN	Q52qr4 glycine max	245	41.5	39.9	825	2	Q7Z7W7_COPCI	Q7z7w7 coprinus ci
173	42	40.4	302	2	Q2X1W5_9GAMM	Q2x1w5 shewanella	246	41	39.4	91	2	Q8W3Q1_9CARY	Q8w3q1 silene arme
174	42	40.4	302	2	Q2ZLG1_SHEPU	Q2zlg1 shewanella	247	41	39.4	138	2	Q4QLY4_HAEI8	Q4qly4 haemophilus
175	42	40.4	319	2	Q567B5_BRARE	Q567b5 brachydanio	248	41	39.4	140	2	Q3VNC7_9CHLB	Q3vnc7 pelodictyon
176	42	40.4	320	2	Q6DIL7_XENTR	Q6dil7 xenopus tro	249	41	39.4	144	2	Q22176_CAEEL	Q22176 caenorhabdi
177	42	40.4	326	1	HFLC_VIBPA	P40606 vibrio para	250	41	39.4	147	2	Q9BH18_GLOPA	Q9bh18 globodera p

251	41	39.4	149	2	Q4QKTS_HAE18	Q4qkt5 haemophilus	324	41	39.4	754	2	Q5H9N1_GLUOX	Q5h9n1 gluconobact
252	41	39.4	151	2	Q2V5D0_NITMU	Q2v5d0 nitrospir	325	41	39.4	765	2	Q4P0C9_USTMA	Q4p0c9 ustilago ma
253	41	39.4	158	2	Q86FK2_HETGL	Q86fk2 heterodera	326	41	39.4	765	2	Q82T03_NITEU	Q82t03 nitroaomona
254	41	39.4	164	2	Q3VIB7_9CHLB	Q3vib7 pelodictyon	327	41	39.4	778	2	Q7RWF9_NEUCR	Q7rwf9 neurospora
255	41	39.4	208	2	Q3UND9_MOUSE	Q3und9 mus musculus	328	41	39.4	785	2	Q9YCC8_AERPE	Q9ycc8 aeropyrum p
256	41	39.4	227	1	LIPB_SYNY3	Q3wib7 pelodictyon	329	41	39.4	839	2	Q3NSP0_SHEFR	Q3nsp0 shewanella
257	41	39.4	229	2	Q4RPX4_TETNG	P74519 syntechocyt	330	41	39.4	848	2	Q3KIB4_PORG1	Q3kib4 porphyromon
258	41	39.4	242	2	Q8TL70_METAC	Q4tpx4 tetraodon n	331	41	39.4	990	2	Q75C89_ASHGO	Q75c89 agbhya goes
259	41	39.4	244	2	Q37KM2_RHOPA	Q8l170 methanosarc	332	41	39.4	997	2	Q4I711_GIBZE	Q4i711 gibberella
260	41	39.4	244	2	Q2J482_RHOPA	Q37km2 rhodospseudo	333	41	39.4	1002	2	Q6CSJ4_KLUUA	Q6csj4 kluyveromyc
261	41	39.4	255	2	Q5L8S6_BACFN	Q2j482 rhodopseudo	334	41	39.4	1119	2	Q3HFQ0_TRIER	Q3hfq0 trichodesmi
262	41	39.4	255	2	Q64NYS_BACFR	Q5l8s6 bacteroides	335	41	39.4	1147	2	Q5YW29_NOCFA	Q5yw29 nocardia fa
263	41	39.4	257	2	Q871L9_NEUCR	Q64nys bacteroides	336	41	39.4	1242	2	Q612B7_CABBR	Q612b7 caenorhabdi
264	41	39.4	260	2	Q75LC8_ORYSA	Q871l9 neurospora	337	41	39.4	1341	2	Q4DXC3_TRYCR	Q4dxc3 trypanosoma
265	41	39.4	262	2	Q69U16_ORYSA	Q75lc8 oryza sativ	338	41	39.4	1351	2	Q4CW02_TRYCR	Q4cw02 trypanosoma
266	41	39.4	262	2	Q58002_SYNY3	Q69u16 oryza sativ	339	41	39.4	1448	2	Q7R0M8_GIALA	Q7r0m8 giardia lam
267	41	39.4	264	1	CAH7_MOUSE	Q58002 syntechocyt	340	41	39.4	1487	2	Q9FH23_ARATH	Q9fh23 arabidopsis
268	41	39.4	267	2	Q349R4_RHOPA	Q9qrq8 mus musculus	341	41	39.4	1478	2	Q3FI51_9BURK	Q3fi51 burkholderi
269	41	39.4	273	2	Q50WB2_ENTHI	Q349r4 rhodopseudo	342	41	39.4	5060	2	O52545_ARYMD	O52545 amycolator
270	41	39.4	276	2	Q7VPD0_HAEDU	Q50wb2 entamoeba h	343	41	39.4	5069	2	O52789_ARYMD	O52789 amycolator
271	41	39.4	281	2	Q58Z98_HAEIN	Q7vpd0 haemophilus	344	41	39.4	6675	2	Q4KCD8_PSEF5	Q4kcd8 pseudomonas
272	41	39.4	300	2	Q8NK77_NECHA	Q58z98 haemophilus	345	41	39.4	142	2	Q8I122_CAEEL	Q8i122 caenorhabdi
273	41	39.4	300	2	Q368C7_9GAMM	Q8nk77 fusarium so	346	40.5	38.9	150	2	Q3BT70_XANC5	Q3bt70 xanthomonas
274	41	39.4	302	2	Q222W6_9GAMM	Q368c7 shewanella	347	40.5	38.9	186	2	O52166_CHICK	O52166 gallus gall
275	41	39.4	302	2	Q35S03_9GAMM	Q222w6 shewanella	348	40.5	38.9	243	2	Q4QA14_LEIMA	Q4qa14 leishmania
276	41	39.4	306	2	Q3RQ22_RALME	Q35s03 shewanella	349	40.5	38.9	335	2	Q4WFF2_ASFPU	Q4wff2 aspergillus
277	41	39.4	312	2	Q7WMN5_BORPE	Q3rq22 ralstonia m	350	40.5	38.9	337	2	Q6K3L1_ORYSA	Q6k3l1 oryza sativ
278	41	39.4	312	2	Q7W9H8_BORPA	Q7wmn5 bordetella	351	40.5	38.9	347	2	O8ELX7_OCEIH	O8elx7 oceanobacil
279	41	39.4	312	2	Q7WH83_BORBR	Q7w9h8 bordetella	352	40.5	38.9	553	2	Q81D71_BACRC	Q81d71 bacillus ce
280	41	39.4	316	2	O5RJX5_XENLA	Q7wh83 bordetella	353	40.5	38.9	792	2	Q7Q986_ANOGA	Q7q986 anopheles g
281	41	39.4	318	2	Q470H3_RALEJ	O5rjx5 xenopus lae	354	40.5	38.9	801	2	Q4Q098_PSE14	Q4q098 pseudomonas
282	41	39.4	333	2	Q3DVA4_STRAG	Q470h3 ralstonia e	355	40.5	38.9	1133	2	Q4HYF1_GIBZE	Q4hyf1 gibberella
283	41	39.4	333	2	Q8DY58_STRAS	Q3dva4 streptococc	356	40	38.5	92	2	O8W3Q2_9CARY	O8w3q2 silene arne
284	41	39.4	346	2	Q8LFD4_ARATH	Q8dy58 streptococc	357	40	38.5	92	2	Q8W3Q7_SILCO	Q8w3q7 silene con
285	41	39.4	346	2	Q3J7Y7_NITOC	Q8lfd4 arabidopsis	358	40	38.5	92	2	O8W3Q9_SILDC	O8w3q9 silene dici
286	41	39.4	346	2	Q41X25_DESHA	Q3j7y7 nitoc	359	40	38.5	92	2	O8W3R9_SILDI	O8w3r9 silene diol
287	41	39.4	350	2	Q7N0M1_PHOLL	Q41x25 deaulfitoba	360	40	38.5	92	2	O8W3S3_SILDI	O8w3s3 silene diol
288	41	39.4	358	2	Q6D819_ERWCT	Q7n0m1 photorhabdu	361	40	38.5	92	2	O8W3S4_SILDI	O8w3s4 silene diol
289	41	39.4	369	2	Q5F4H4_NAIZE	Q6d819 erwinia car	362	40	38.5	99	2	O8U4W7_AGR75	O8u4w7 agrobacteri
290	41	39.4	370	2	Q4T930_TETNG	Q5f4h4 zea mays (m	363	40	38.5	116	2	O65048_PICWA	O65048 picea maria
291	41	39.4	379	2	Q4UC90_THEAN	Q4t930 tetraodon n	364	40	38.5	141	2	O65048_PICWA	O65048 picea maria
292	41	39.4	382	2	Q3E8P7_ARATH	Q4uc90 theileria a	365	40	38.5	155	2	Q44P62_CHLLI	Q44p62 chlorobium
293	41	39.4	386	2	Q391Y5_BURS3	Q3e8p7 arabidopsis	366	40	38.5	167	2	Q2JIJ6_9CYAN	Q2ji6 cyanobacter
294	41	39.4	422	2	Q4N003_THEPA	Q391y5 burkholderi	367	40	38.5	171	2	Q5U7L6_CABRE	Q5u7l6 caenorhabdi
295	41	39.4	426	2	Q4HZC4_GIBZE	Q4n003 theileria p	368	40	38.5	178	2	Q8Y981_LISMO	Q8y981 listeria mo
296	41	39.4	431	1	Y4BN_RHISN	Q4hzc4 gibberella	369	40	38.5	178	2	Q8Y981_LISMO	Q8y981 listeria mo
297	41	39.4	444	2	Q9X2D2_THEMEA	P55381 rhizobium s	370	40	38.5	188	2	Q3CMY8_ALTAT	Q3cm8 pseudocalter
298	41	39.4	453	2	Q8GLP5_AERHY	Q9x2d2 thermotoga	371	40	38.5	202	2	Q3M698_ANAVT	Q3m698 anabaena va
299	41	39.4	454	2	Q2JU25_9CIAN	Q8glp5 aeromonas h	372	40	38.5	221	1	LIFB_ANASP	Lifb anabaena sp
300	41	39.4	457	2	Q94LG3_ORYSA	Q2ju25 cyanobacter	373	40	38.5	225	2	Q2KJ700_BOVIN	Q2kj700 bos taurus
301	41	39.4	471	2	Q9FWQ9_ARATH	Q94lg3 oryza sativ	374	40	38.5	235	2	Q8K793_STRP3	Q8k793 streptococc
302	41	39.4	471	2	Q8BGL4_PSEPK	Q9fwq9 arabidopsis	375	40	38.5	243	2	O57E28_BRUAB	O57e28 brucella ab
303	41	39.4	504	2	Q95XE7_CAEEL	Q8bgl4 pseudomonas	376	40	38.5	243	2	O8G2G0_BRUSU	O8g2g0 brucella su
304	41	39.4	514	2	Q4Q181_LEIMA	Q95xr7 caenorhabdi	377	40	38.5	243	2	O8YFF7_BRUME	O8yff7 brucella me
305	41	39.4	514	2	Q6K1E0_MYCMO	Q4qi81 leishmania	378	40	38.5	246	2	Q2YM83_BRUAX	Q2ym83 brucella ab
306	41	39.4	516	2	Q4IHR1_GIBZE	Q6k1e0 mycoplasma	379	40	38.5	246	2	Q2NM83_XANOR	Q2nm83 xanthomonas
307	41	39.4	517	2	O8AW13_BRARE	Q4ihr1 gibberella	380	40	38.5	253	2	Q424Y0_DESHA	Q424y0 desulfitoba
308	41	39.4	529	2	O8L7U6_ARATH	O8aw13 brachydanio	381	40	38.5	262	2	O9VIX5_DROME	O9vix5 drosophila
309	41	39.4	536	1	HPSE3_ARATH	O8l7u6 arabidopsis	382	40	38.5	262	2	Q3X3D7_9ACTN	Q3x3d7 rubrobacter
310	41	39.4	537	2	Q6CR42_KLUUA	Q9fzpl arabidopsis	383	40	38.5	265	2	Q5GYV5_XANOR	Q5gyv5 xanthomonas
311	41	39.4	551	2	Q2UUI6_ASPOR	Q6cr42 kluyveromyc	384	40	38.5	269	2	Q8CUW7_OCEIH	Q8cuw7 oceanobacil
312	41	39.4	570	2	Q3QJD8_9GAMM	Q2uui6 aspergillus	385	40	38.5	278	2	Q3SK37_9BRAD	Q3sk37 bradyrhizob
313	41	39.4	597	1	GP63_LEIAM	Q3qjd8 shewanella	386	40	38.5	286	1	DAPA2_CLOAB	Dapa2 clostridium
314	41	39.4	643	2	Q9J0H4_9VIRU	Q2t673 leishmania	387	40	38.5	287	2	O9VIX6_DROME	O9vix6 drosophila
315	41	39.4	660	2	Q66925_AQUAE	Q9j0h4 zygoecchar	388	40	38.5	291	2	Q2UAS9_ASPOR	Q2uas9 aspergillus
316	41	39.4	678	2	Q3WTC8_RRHIZ	Q66925 aquifex aeo	389	40	38.5	297	2	Q81DG9_BACCR	Q81dg9 bacillus ce
317	41	39.4	681	2	Q3DXH8_CHLUA	Q3wtc8 mesorhizobi	390	40	38.5	298	2	Q3EQV1_BACTI	Q3eqv1 bacillus th
318	41	39.4	682	2	Q7W592_BORPA	Q3dxh8 chloroflexu	391	40	38.5	299	2	Q4MHL6_BACCE	Q4mhl6 bacillus ce
319	41	39.4	682	2	Q7MCS3_BORBR	Q7w592 bordetella	392	40	38.5	299	2	O63BB1_BACCC	O63bb1 bacillus th
320	41	39.4	687	2	Q3ILX0_NATPD	Q7mcs3 bordetella	393	40	38.5	299	2	O6HIP3_BACCH	O6hip3 bacillus th
321	41	39.4	690	2	Q7VZR8_BORPE	Q3ily0 natronomona	394	40	38.5	299	2	Q737Z6_BACCI	Q737z6 bacillus ce
322	41	39.4	718	2	Q46TM0_RALEJ	Q7vzr8 bordetella	395	40	38.5	300	2	Q2RT98_RHORU	Q2rt98 rhodospiril
323	41	39.4	730	2	Q6WB73_ALCFA	Q46tm0 ralstonia e	396	40	38.5	303	2	Q46NY0_RALEJ	Q46ny0 ralstonia e

397	40	38.5	305	2	086906	98PHN	086906	sphingomona	470	40	38.5	481	2	Q5NH7	FRATT	Q5nhn7	francisella
398	40	38.5	308	2	Q25B4	9GAMM	Q25B4	shewanella	471	40	38.5	495	2	Q3JD49	NITOC	Q3jd49	nitrosococc
399	40	38.5	308	2	Q36012	shewanella	Q36012	shewanella	472	40	38.5	497	2	Q5ADH5	CANAL	Q5adh5	canida alb
400	40	38.5	308	2	Q36D22	9GAMM	Q36d22	shewanella	473	40	38.5	505	2	Q45905	COXBU	Q45905	coxiella bu
401	40	38.5	310	2	Q9X7X2	STRCO	Q9x7x2	streptomyc	474	40	38.5	512	2	Q8VUT0	PARDE	Q8vut0	paracoccus
402	40	38.5	311	1	GLUQ	NOCFA	Q5z380	nocardia fa	475	40	38.5	526	2	Q52883	COXBU	Q52883	coxiella bu
403	40	38.5	315	2	Q3NR25	SHEFR	Q3nr25	shewanella	476	40	38.5	526	2	Q9X626	COXBU	Q9x626	coxiella bu
404	40	38.5	326	2	Q9VW53	DROME	Q9vw53	drosophila	477	40	38.5	528	2	Q53949	COXBU	Q53949	coxiella bu
405	40	38.5	328	2	Q4QN35	HAB18	Q4qn35	haemophilus	478	40	38.5	529	2	Q53HS8	HUMAN	Q53hs8	homo sapien
406	40	38.5	333	2	Q3D2B9	STRAG	Q3d2e9	streptococc	479	40	38.5	540	2	Q827B4	STRAW	Q827b4	streptomyc
407	40	38.5	333	2	Q3D8R7	STRAG	Q3d8r7	streptococc	480	40	38.5	546	2	Q5AD44	CANAL	Q5ad44	canida alb
408	40	38.5	333	2	Q3DMN2	STRAG	Q3dmn2	streptococc	481	40	38.5	550	2	Q5AD47	CANAL	Q5ad47	canida alb
409	40	38.5	333	2	Q8E4D6	STRA3	Q8e4d6	streptococc	482	40	38.5	573	2	Q6A3Q2	CABEL	Q6a3q2	caenorhabdi
410	40	38.5	338	2	Q76655	CABEL	Q76655	caenorhabdi	483	40	38.5	576	1	SBP	CABEL	Q21950	caenorhabdi
411	40	38.5	346	2	Q66DS4	YERPS	Q66ds4	yersinia ps	484	40	38.5	580	2	Q8NSW4	CORGL	Q8nsw4	corynebacte
412	40	38.5	346	2	Q82C75	YERPE	Q82c75	yersinia pe	485	40	38.5	583	2	Q3P6N6	9GAMM	Q3p6n6	shewanella
413	40	38.5	349	2	Q68067	RHOCA	Q68067	rhodobacter	486	40	38.5	595	2	Q7ZX69	XENLA	Q7zx69	shenopus lae
414	40	38.5	356	2	Q8G1B8	9MYCO	Q8g1b8	mycobacteri	487	40	38.5	606	2	Q9N3L4	CABEL	Q9n3l2	caenorhabdi
415	40	38.5	356	2	Q8GR6C	BACSU	Q8grc6	bacillus su	488	40	38.5	619	2	Q4K3N4	PSEF5	Q4k3n4	pseudomonas
416	40	38.5	363	2	Q2KA13	RHIET	Q2ka13	rhizobium e	489	40	38.5	623	1	MCE1	CABEL	Q17607	caenorhabdi
417	40	38.5	363	2	Q70P84	9TOMB	Q70p84	johnsongras	490	40	38.5	635	2	Q6M7K0	CORGL	Q6m7k0	corynebacte
418	40	38.5	365	1	SOXB	RHOGG	P54997	rhodococcus	491	40	38.5	648	2	Q6BV31	DEBHA	Q6bv31	debaryomyce
419	40	38.5	365	2	Q4PPA2	BREBE	Q4ppa2	brevibacill	492	40	38.5	660	2	Q5E832	VIBF1	Q5e832	vibrio fisc
420	40	38.5	365	2	Q4PPA8	ACIDE	Q4ppa8	acidovorax	493	40	38.5	672	2	Q6CWC7	KLULA	Q6cwc7	kluyveromyc
421	40	38.5	365	2	Q56GA0	9RHI2	Q56ga0	agrobacteri	494	40	38.5	684	2	Q4P919	USTMA	Q4p919	ustilago ma
422	40	38.5	365	2	Q5U773	9NOCA	Q5u773	rhodococcus	495	40	38.5	706	2	Q6MK05	BDEHA	Q6mk05	bdellovibri
423	40	38.5	365	2	Q64F41	NOCGL	Q64f41	nocardia gl	496	40	38.5	721	2	Q2RS04	RHOER	Q2rs04	rhodospiril
424	40	38.5	365	2	Q64F44	RHOER	Q64f44	rhodococcus	497	40	38.5	728	2	Q4N0A0	THEPA	Q4n0a0	theileria p
425	40	38.5	365	2	Q64F47	9ACTO	Q64f47	gordonia al	498	40	38.5	742	1	DHET	AGEAC	P18278	acetobacter
426	40	38.5	365	2	Q6WE14	RHOER	Q6we14	rhodococcus	499	40	38.5	742	2	Q53362	ACEPA	Q53362	acetobacter
427	40	38.5	365	2	Q6WEB1	9ACTO	Q6web1	gordonia al	500	40	38.5	755	2	Q615F2	CAEBR	Q615f2	caenorhabdi
428	40	38.5	365	2	Q6WNP2	9NOCA	Q6wnp2	rhodococcus	501	40	38.5	757	1	DHET	GLUOX	Q05542	gluonobact
429	40	38.5	365	2	Q70WT1	9NOCA	Q70wt1	rhodococcus	502	40	38.5	762	2	Q22168	ARATH	Q22168	arabidopsis
430	40	38.5	376	2	Q63006	BACCZ	Q630q6	bacillus ce	503	40	38.5	765	2	Q8RY37	ARATH	Q8ry37	arabidopsis
431	40	38.5	376	2	Q6HAU1	BACHK	Q6hau1	bacillus th	504	40	38.5	853	2	Q3AR17	CHLCH	Q3ar17	chlorobium
432	40	38.5	376	2	Q81JV7	BACAN	Q81jv7	bacillus an	505	40	38.5	966	2	Q9LNT7	ARATH	Q9lnt7	arabidopsis
433	40	38.5	381	1	SEPP1	HUMAN	P49908	homo sapien	506	40	38.5	996	2	Q4WEW1	ASPFU	Q4wew1	aspergillus
434	40	38.5	381	1	SEPP1	PONPY	Q5r8w9	pongo pygma	507	40	38.5	1131	2	Q8JKN7	9VIRU	Q8jkn7	heliothis z
435	40	38.5	382	2	Q2IA49	KARMI	Q2ia49	karlodinium	508	40	38.5	1202	2	Q8JKN7	9VIRU	Q8jkn7	heliothis z
436	40	38.5	384	2	Q6P2U7	BRARE	Q6p2u7	brachydanio	509	40	38.5	1259	2	Q5LR00	SILPO	Q5lr00	silicibacte
437	40	38.5	386	2	Q83BP5	COXBU	Q83bp5	coxiella bu	510	40	38.5	1301	2	Q96Y03	SULTO	Q96y03	sulfolobus
438	40	38.5	386	2	Q7ZVX1	BRARE	Q7zvxl	brachydanio	511	40	38.5	1364	2	Q4B4P7	9BURK	Q4b4p7	polaromonas
439	40	38.5	388	2	Q454J1	9BURK	Q454j1	burkholderi	512	40	38.5	1451	2	Q5CNE9	CRYHO	Q5cne9	cryptospori
440	40	38.5	388	2	Q4LNE3	9BURK	Q4lne3	burkholderi	513	40	38.5	1455	2	Q4IVL8	AZOV1	Q4ivl8	azotobacter
441	40	38.5	393	2	Q732P1	BACCI	Q732p1	bacillus ce	514	40	38.5	1498	2	Q8IDP2	PLAF7	Q8idp2	plasmodium
442	40	38.5	396	1	XJR1	YEAST	P46992	saccharomyc	515	40	38.5	1595	2	Q2U5K9	ASPOR	Q2u5k9	aspergillus
443	40	38.5	399	2	Q30S48	THIDN	Q30s48	thiomicrosp	516	40	38.5	1674	2	Q5CZ08	CRYPV	Q5cz08	cryptospori
444	40	38.5	403	2	Q3W3L9	9ACTO	Q3w3l9	frankia sp.	517	40	38.5	2081	2	Q2UR04	9ACTO	Q2ur04	frankia sp.
445	40	38.5	404	2	Q8LD23	ARATH	Q8ld23	arabidopsis	518	40	38.5	2456	2	Q2UNE1	ASFOR	Q2une1	aspergillus
446	40	38.5	405	2	Q4E144	TRYCR	Q4e144	trypanosoma	519	40	38.5	2638	2	Q3GC27	9FIRM	Q3gc27	syntrophomo
447	40	38.5	418	2	Q7XRK4	ORYSA	Q7xrkl	oryza sativ	520	40	38.5	3081	2	Q2WNY2	GLOBE	Q2wny2	clostridium
448	40	38.5	419	2	Q01320	CABEL	Q01320	caenorhabdi	521	40	38.5	4007	2	Q2J2W7	RHOPA	Q2j2w7	rhodopseudo
449	40	38.5	420	2	Q2YD66	BOVIN	Q2ydg6	bos taurus	522	40	38.5	6256	2	Q4U446	POLCB	Q4u446	polyangium
450	40	38.5	426	2	Q4AFJ9	9CHLB	Q4afj9	chlorobium	523	40	38.5	8417	2	Q2N3T0	POLCB	Q2n3t0	polyangium
451	40	38.5	427	2	Q3QSS9	9RHOB	Q3qss9	silicibacte	524	39.5	38.0	135	2	Q5P440	AZOSE	Q5p440	azocarcus sp
452	40	38.5	432	2	Q8KFU4	CHLTE	Q8kfu4	chlorobium	525	39.5	38.0	163	2	Q7BKF3	PRB01	Q7bkf3	gamma-prote
453	40	38.5	433	2	Q68WJ9	RICTY	Q68wj9	rickettsia	526	39.5	38.0	168	2	Q3B264	XANCS	Q3b264	xanthomonas
454	40	38.5	433	2	Q9ZD14	RICPR	Q9zdl4	rickettsia	527	39.5	38.0	195	2	Q3J0U9	RHOS4	Q3j0u9	rhodobacter
455	40	38.5	438	2	Q92JG9	RICCN	Q92jg9	rickettsia	528	39.5	38.0	223	2	Q4HG06	CAMCO	Q4hg06	campylobact
456	40	38.5	439	2	Q3J0J6	9GAMM	Q3j0j6	alkalilimni	529	39.5	38.0	223	2	Q9FIE6	CAMJE	Q9fie6	campylobact
457	40	38.5	443	2	Q5KJ41	CRYNE	Q5kj41	cryptococcu	530	39.5	38.0	223	2	Q5HWB2	CAMJR	Q5hwb2	campylobact
458	40	38.5	448	2	Q846U7	9BACL	Q846u7	brevibacill	531	39.5	38.0	246	2	P70884	BUTFI	P70884	butyrivibri
459	40	38.5	453	2	Q6CT05	KLULA	Q6ctq5	kluyveromyc	532	39.5	38.0	256	2	Q84B54	9BACT	Q84b54	uncultured
460	40	38.5	461	2	Q57WK7	9TRYP	Q57wk7	trypanosoma	533	39.5	38.0	289	1	DAFF	STRCT	Q69969	streptomyc
461	40	38.5	462	2	Q7XG16	ORYSA	Q7xg16	oryza sativ	534	39.5	38.0	289	2	Q82KD0	STRAW	Q82kd0	streptomyc
462	40	38.5	463	1	HSLU	BACAN	Q81wk6	bacillus an	535	39.5	38.0	383	2	Q6GH10	STAAH	Q6gh10	staphylococ
463	40	38.5	463	1	HSLU	BACR	Q819x7	bacillus ce	536	39.5	38.0	383	2	Q8AAS3	BACTN	Q8aaas3	bacteroides
464	40	38.5	463	1	HSLU	BACCZ	Q636j7	bacillus ce	537	39.5	38.0	387	2	Q4K7B0	PSEF5	Q4k7b0	pseudomonas
465	40	38.5	463	1	HSLU	BACHK	Q6ney6	bacillus th	538	39.5	38.0	388	2	Q46524	FELCA	Q46524	felis silve
466	40	38.5	463	2	Q4MGV4	BACCCE	Q4mgv4	bacillus ce	539	39.5	38.0	433	2	Q93G99	FUSNU	Q93g99	fusobacteri
467	40	38.5	474	2	Q4MTW4	BACCCE	Q4mtw4	bacillus ce	540	39.5	38.0	454	2	Q41GR9	9BACI	Q41gr9	exiguobacte
468	40	38.5	481	2	Q55T99	CRYNE	Q55t99	cryptococcu	541	39.5	38.0	480	2	Q8DLQ3	SYNEL	Q8dlq3	synechococc
469	40	38.5	481	2	Q5KJ42	CRYNE	Q5kj42	cryptococcu	542	39.5	38.0	481	2	Q9U8C5	CABEL	Q9u8c5	caenorhabdi

543	39.5	38.0	555	2	Q5B9I6	EMENI	Q5B9I6	aspergillus	616	39	37.5	248	1	ECX1_THEAC	Q9hip2	thermoplasm
544	39.5	38.0	575	2	Q5CJ83	CRYHO	Q5CJ83	cryptospori	617	39	37.5	248	1	ECX1_THEVO	Q97b25	thermoplasm
545	39.5	38.0	576	2	Q4RP41	9HEPC	Q4RP41	hepatitis c	618	39	37.5	248	2	Q4SVI2_TETNG	Q4sv12	tetradodon n
546	39.5	38.0	576	2	Q4KP46	9HEPC	Q4KP46	hepatitis c	619	39	37.5	250	2	Q36WP8_RHOPA	Q36wp8	rhodopsendo
547	39.5	38.0	576	2	Q4KP47	9HEPC	Q4KP47	hepatitis c	620	39	37.5	254	2	Q36QJ6_9RHOB	Q36qj6	silicibacte
548	39.5	38.0	591	2	Q3Q061	9GAMM	Q3Q061	psychrobact	621	39	37.5	254	2	Q6DJ72_XENTH	Q6dj72	xenopus tro
549	39.5	38.0	591	2	Q4FU62	PSVAR	Q4FU62	psychrobact	622	39	37.5	258	2	Q2RHC3_MOOTH	Q2rhc3	moorella th
550	39.5	38.0	591	2	Q3P428	9GAMM	Q3P428	shewanella	623	39	37.5	260	2	Q2TYD4_ASPOR	Q2tyd4	aspergillus
551	39.5	38.0	639	2	Q2WB52	MAGSA	Q2WB52	magnetospir	624	39	37.5	263	2	Q2NND7_XANOR	Q2nzd7	xanthomonas
552	39.5	38.0	840	2	Q6CHB0	YARLI	Q6CHB0	yarrowia li	625	39	37.5	264	2	Q2IIC9_PSEAE	Q2iic9	pseudomonas
553	39.5	38.0	972	2	Q740V3	MYCPA	Q740V3	mycobacteri	626	39	37.5	265	2	Q3B6B9_PSELJD	Q3b6b9	peloidictyon
554	39.5	38.0	974	2	Q9XCF6	MYCAV	Q9XCF6	mycobacteri	627	39	37.5	266	2	Q2XEF7_PSEBU	Q2xef7	pseudomonas
555	39.5	38.0	1012	2	Q3SIU2	THIDA	Q3SIU2	thiobacilli	628	39	37.5	266	2	Q88RB2_PSEPK	Q88rb2	pseudomonas
556	39.5	38.0	1039	2	Q3RSU0	RALME	Q3RSU0	raletonia m	629	39	37.5	267	2	Q9S3Z3_MACFA	Q9s3z3	macaca fasc
557	39.5	38.0	1135	2	Q5E6L9	VIBFI	Q5E6L9	vibri	630	39	37.5	267	2	Q40JBO_EHRCH	Q40jbo	ehrlichia c
558	39.5	38.0	3483	2	Q4E5K4	TRYCR	Q4E5K4	trypanosoma	631	39	37.5	271	2	Q63AP4_BACCC	Q63ap4	bacillus ce
559	39.5	38.0	3488	2	Q4E5L0	TRYCR	Q4E5L0	trypanosoma	632	39	37.5	271	2	Q6F9B3_ACTAD	Q6f9b3	acinetobact
560	39	37.5	38	2	Q9S565	DROME	Q9S565	drosophila	633	39	37.5	271	2	Q6HI32_BACHK	Q6hi32	bacillus th
561	39	37.5	42	2	Q73EW2	BACCL	Q73EW2	bacillus ce	634	39	37.5	271	2	Q81PV2_BACAN	Q81pv2	bacillus an
562	39	37.5	63	2	Q4XDES	PLACH	Q4XDES	plasmodium	635	39	37.5	285	2	Q2WQL1_CLOBE	Q2wql1	clostridium
563	39	37.5	92	2	Q8W3P9	VISVU	Q8W3P9	viscaria vu	636	39	37.5	289	2	Q2XBM9_PSEPU	Q2xbm9	pseudomonas
564	39	37.5	92	2	Q8W3Q0	VISVU	Q8W3Q0	viscaria vu	637	39	37.5	289	2	Q88DD6_PSEPK	Q88dd6	pseudomonas
565	39	37.5	92	2	Q8W3R7	SILLA	Q8W3R7	silene lati	638	39	37.5	292	2	Q5NKS3_SORBI	Q5nks3	sorghum bic
566	39	37.5	92	2	Q8W3T0	9CARY	Q8W3T0	lychnis cor	639	39	37.5	294	2	Q4WA05_ASPFU	Q4wa05	aspergillus
567	39	37.5	92	2	Q8W3T1	9CARY	Q8W3T1	lychnis cor	640	39	37.5	295	2	Q3X6K3_METFL	Q3x6k3	methylobaci
568	39	37.5	97	2	Q41JW2	GIBBE	Q41JW2	gibberella	641	39	37.5	296	1	Y1523_HAEIN	Y41233	haemophilus
569	39	37.5	103	2	Q5YFC6	9VIRU	Q5YFC6	singapore g	642	39	37.5	296	2	Q3RUM0_RALME	Q3rum0	raletonia m
570	39	37.5	106	2	Q75703	HUMAN	Q75703	homo sapien	643	39	37.5	301	2	Q5J964_PYRHO	Q5je23	pyrococcus
571	39	37.5	118	2	Q5TYI6	ANOGA	Q5TYI6	anopheles g	644	39	37.5	303	2	Q5J923_PYRKO	Q5je23	pyrococcus
572	39	37.5	128	2	Q3IZ54	RHOSA	Q3IZ54	rhodobacter	645	39	37.5	304	2	Q00867_NECHA	Q00867	nectria hae
573	39	37.5	132	2	Q4IKL5	PLABE	Q4IKL5	plasmodium	646	39	37.5	305	2	Q3MHU9_HUMAN	Q3mhu9	homo sapien
574	39	37.5	143	2	Q22032	CABEL	Q22032	caenorhabdi	647	39	37.5	314	2	Q5T5I5_HUMAN	Q5t5i5	homo sapien
575	39	37.5	143	2	Q61K63	CAEBR	Q61K63	caenorhabdi	648	39	37.5	318	2	Q96LS2_HUMAN	Q96ls2	homo sapien
576	39	37.5	149	2	Q442N9	SOLUS	Q442N9	solibacter	649	39	37.5	319	2	Q8LB51_ARATH	Q8lb51	arabidopsais
577	39	37.5	156	2	Q6TUJ3	ASCUS	Q6TUJ3	ascaris suu	650	39	37.5	319	2	Q9M2J6_ARATH	Q9m2j6	arabidopsais
578	39	37.5	156	2	Q6V7X0	ASCUS	Q6V7X0	ascaris suu	651	39	37.5	324	2	Q9HX84_PSEAE	Q9hx84	pseudomonas
579	39	37.5	158	2	Q5J1H7	PYRKO	Q5J1H7	pyrococcus	652	39	37.5	327	2	Q4UI51_THEAN	Q4uis1	theileria a
580	39	37.5	158	2	Q665E0	YERPS	Q665E0	yersinia ps	653	39	37.5	329	2	Q59480_PYRHO	Q59480	pyrococcus
581	39	37.5	158	2	Q82AY0	YERPE	Q82AY0	yersinia pe	654	39	37.5	330	2	Q8SW16_ENCCU	Q8sw16	encephalito
582	39	37.5	164	2	Q5T5I4	HUMAN	Q5T5I4	homo sapien	655	39	37.5	333	2	Q31522_POBRE	Q31522	poecilia re
583	39	37.5	166	2	Q413B5	GIBZE	Q413B5	gibberella	656	39	37.5	334	2	Q3X0H3_9ACTN	Q3x0h3	rubrobacter
584	39	37.5	166	2	Q8SNF8	XIPHE	Q8SNF8	xiphophorus	657	39	37.5	339	1	FMRP_DROVI	P41876	drosophila
585	39	37.5	171	2	Q6SHET	9ABPA	Q6shet	vibrio para	658	39	37.5	345	2	Q5LNP7_SILPO	Q5lnp7	silicibacte
586	39	37.5	173	2	Q6SHET	9ABPA	Q6shet	uncultured	659	39	37.5	350	2	Q2K4B2_RHIEF	Q2k4b2	rhizobium e
587	39	37.5	186	2	Q9A6R0	CAUCR	Q9A6R0	caulobacter	660	39	37.5	361	2	Q7RD19_PLAYO	Q7rd19	plasmodium
588	39	37.5	188	2	Q2TB87	9CYAN	Q2TB87	cyanobacter	661	39	37.5	362	2	Q7PVC2_ANOGA	Q7pvc2	anopheles g
589	39	37.5	190	2	Q5T5I3	HUMAN	Q5T5I3	homo sapien	662	39	37.5	362	2	Q4AFG2_9CHUB	Q4afg2	chlorobium
590	39	37.5	191	2	Q9SNI6	ORYSA	Q9SNI6	oryza sativ	663	39	37.5	362	2	Q3FP95_9BURK	Q3fp95	rhodoferrax
591	39	37.5	192	1	MOBA_HAEIN		P44899	haemophilus	664	39	37.5	365	2	Q9HHX5_HALSA	Q9hhx5	halobacteri
592	39	37.5	192	2	Q4QM57	HAEI8	Q4qm57	haemophilus	665	39	37.5	369	2	Q5P5Q9_AZOSE	Q5p5q9	methanosarc
593	39	37.5	201	2	Q4AFU2	9CHLB	Q4afu2	chlorobium	666	39	37.5	374	2	Q466B0_METBA	Q466b0	methanosarc
594	39	37.5	202	2	Q6AJME	DSRPS	Q6ajme	desulfotale	667	39	37.5	374	2	Q8TQP2_METAC	Q8tqp2	methanosarc
595	39	37.5	205	2	Q9ON07	CAPHI	Q9on07	capra hircu	668	39	37.5	375	2	Q97BR9_THEVO	Q97br9	thermoplasm
596	39	37.5	208	2	Q7LAD9	HUMAN	Q7lad9	homo sapien	669	39	37.5	376	2	Q4MIE7_BACCE	Q4mie7	bacillus ce
597	39	37.5	208	2	Q5L7T2	BACFN	Q5l7t2	bacteroides	670	39	37.5	376	2	Q98AF3_RHIL0	Q98af3	rhizobium l
598	39	37.5	208	2	Q64N03	BACER	Q64n03	bacteroides	671	39	37.5	376	2	Q72XA5_BACC1	Q72xa5	bacillus ce
599	39	37.5	209	2	Q7R565	GIALA	Q7r565	giardia lam	672	39	37.5	380	1	SEPP1_MOUSE	Q80uf3	m adult mal
600	39	37.5	209	2	Q8A573	BACTN	Q8a573	bacteroides	673	39	37.5	380	2	Q3TJ31_MOUSE	Q3tj31	mus musculu
601	39	37.5	210	2	Q2UI41	ASPOR	Q2ui41	aspergillus	674	39	37.5	380	2	Q3TXG8_MOUSE	Q3txg8	mus musculu
602	39	37.5	217	2	Q2JFL2	9ACTO	Q2jfl2	frankia sp.	675	39	37.5	380	2	Q6PKE7_MOUSE	Q6pke7	mus musculu
603	39	37.5	222	2	Q2S121	9SPHI	Q2s121	salinibacte	676	39	37.5	380	2	Q8OT08_MOUSE	Q8ot08	mus musculu
604	39	37.5	224	1	CBIO_PHOLL		Q7n0n3	photorhabdu	677	39	37.5	380	2	Q80UF3_MOUSE	Q80uf3	m adult mal
605	39	37.5	229	2	Q2WRU6	CLOBE	Q2wru6	clostridium	678	39	37.5	381	2	Q4YIC1_PLACH	Q4yic1	plasmodium
606	39	37.5	230	2	Q38IH8	9BACT	Q38ih8	uncultured	679	39	37.5	381	2	Q4Z238_PLABE	Q4z238	plasmodium
607	39	37.5	231	2	Q3VR24	PROAE	Q3vr24	prothecoch	680	39	37.5	381	2	Q3EZ11_BACTI	Q3ez11	bacillus th
608	39	37.5	241	2	Q3J013	RHOSA	Q3j013	rhodobacter	681	39	37.5	381	2	Q814S8_BACCR	Q814s8	bacillus ce
609	39	37.5	241	2	Q53112	RHOSH	Q53112	rhodobacter	682	39	37.5	384	2	Q61CX6_CABER	Q61cx6	caenorhabdi
610	39	37.5	243	2	Q8KJP0	RHIL0	Q8kjp0	rhizobium l	683	39	37.5	384	2	Q7VEA2_PROWA	Q7vea2	prochloroco
611	39	37.5	243	2	Q5S576	RHIME	Q5s576	rhizobium m	684	39	37.5	385	1	SEPP1_RAT	P25236	rattus norv
612	39	37.5	243	2	Q8UF63	AGRTS	Q8uf63	agrobacteri	685	39	37.5	385	2	Q7V5C4_PROWM	Q7v5c4	prochloroco
613	39	37.5	243	2	Q92ZN9	RHIME	Q92zn9	rhizobium m	686	39	37.5	386	2	Q9N2H6_BOVIN	Q9n2h6	bos taurus
614	39	37.5	243	2	Q989I1	RHIL0	Q989i1	rhizobium l	687	39	37.5	387	2	Q6FIC1_HUMAN	Q6fic1	homo sapien
615	39	37.5	247	2	Q3VGT7	9SPHN	Q3vgt7	sphingopyxi	688	39	37.5	387	2	Q9H0K8_HUMAN	Q9h0k8	homo sapien

689	39	37.5	389	1	RL3_DEBHA	Q6bxs6 debaryomyce	762	39	37.5	510	2	Q7XIE6_ORYSA	Q7xle6 oryza sativ
690	39	37.5	389	2	Q43KD2_9CHLB	Q43kd2 chlorobium	763	39	37.5	510	2	Q3AUM4_SYNS9	Q3aum4 synchococc
691	39	37.5	392	2	Q4XWD6_PLACH	Q4xwd6 plasmoch	764	39	37.5	512	2	Q9NH22_DROME	Q9nhn2 drosophila
692	39	37.5	392	2	Q47GN4_DECAR	Q47gn4 dechloromon	765	39	37.5	517	2	Q30269_ARCFU	Q30269 archaeoglob
693	39	37.5	393	2	Q43136_PLABE	Q43136 plasmodium	766	39	37.5	528	2	Q9Y7C8_ASPTU	Q9y7c8 aspergillus
694	39	37.5	394	2	Q6P6B3_HUMAN	Q6p6b3 homo sapien	767	39	37.5	528	2	Q43LM9_SOLUS	Q43lm9 solibacter
695	39	37.5	394	2	Q86X24_HUMAN	Q86x24 homo sapien	768	39	37.5	534	2	Q54H08_DICDI	Q54h08 dicystosteli
696	39	37.5	394	2	Q4R8B9_MACFA	Q4r8b9 macaca fasc	769	39	37.5	540	2	Q6EAX3_9CALI	Q6eax3 norovirus 9
697	39	37.5	394	2	Q9JGK4_MEHVI	Q9jgk4 meleagrid h	770	39	37.5	546	2	Q57V88_9TRYP	Q57v88 trypanosoma
698	39	37.5	396	2	Q5E4P6_VIBF1	Q5e4p6 vibrio fisc	771	39	37.5	546	2	Q57V89_9TRYP	Q57v89 trypanosoma
699	39	37.5	401	2	Q810S6_DROME	Q810s6 drosophila	772	39	37.5	554	2	Q68FP7_RAT	Q68fp7 rattus norv
700	39	37.5	402	1	SEPP1_BOVIN	P49907 bos taurus	773	39	37.5	558	2	Q55NU4_CRYNE	Q55nu4 cryptococcu
701	39	37.5	402	2	Q5H9X4_ORISA	Q5h9x4 oryza sativ	774	39	37.5	558	2	Q5KEL9_CRYNE	Q5kel9 cryptococcu
702	39	37.5	403	2	Q59KD7_CANAL	Q59kd7 candida alb	775	39	37.5	571	2	Q6BUP5_DEBHA	Q6bup5 debaryomyce
703	39	37.5	404	1	YJ9N_YEAST	P47177 saccharomyc	776	39	37.5	574	2	Q4RJJ2_TETNG	Q4rjj2 tetraodon n
704	39	37.5	408	2	Q61Z13_CAEBR	Q61zi3 caenorhabdi	777	39	37.5	576	2	Q85K17_CHLRE	Q85k17 chlamydomon
705	39	37.5	408	2	Q7D7T7_MYCTU	Q7d7t7 mycobacteri	778	39	37.5	580	2	Q9KBV1_BACHD	Q9kbv1 bacillus ha
706	39	37.5	414	2	Q5OE95_STRFL	Q5oe95 streptomyce	779	39	37.5	581	2	Q9XEK4_BRANA	Q9xek4 brassica na
707	39	37.5	415	2	Q478W4_DECAR	Q478w4 dechloromon	780	39	37.5	585	2	Q4CFP8_CLOTM	Q4cfp8 clostridium
708	39	37.5	417	2	Q9VPG5_DROME	Q9vp95 drosophila	781	39	37.5	586	2	Q61TK3_CAEBR	Q61tk3 caenorhabdi
709	39	37.5	418	2	Q5CFR9_CRYHO	Q5cfr9 cryptospori	782	39	37.5	595	2	Q4D411_TRYCR	Q4d411 trypanosoma
710	39	37.5	418	2	Q5CQ74_CRYPV	Q5cq74 cryptospori	783	39	37.5	595	2	Q4DBE2_TRYCR	Q4dbe2 trypanosoma
711	39	37.5	419	2	Q9VPG4_DROME	Q9vp94 drosophila	784	39	37.5	595	2	Q7QBK9_ANOGA	Q7qbk9 anopheles g
712	39	37.5	422	2	Q07730_MYCTU	Q07730 mycobacteri	785	39	37.5	597	2	Q3CGW0_THERET	Q3cgw0 thermoaer
713	39	37.5	422	2	Q7TZB7_MYCBO	Q7tzb7 mycobacteri	786	39	37.5	606	2	Q2XIX2_PSEPU	Q2xix2 pseudomonas
714	39	37.5	424	2	Q6FP92_CANGA	Q6fp92 candida gla	787	39	37.5	606	2	Q88IA2_PSEPK	Q88ia2 pseudomonas
715	39	37.5	424	2	Q2KYH0_BORAV	Q2kyh0 bordetella	788	39	37.5	619	2	Q4CN95_TRYCR	Q4cn95 trypanosoma
716	39	37.5	424	2	Q7VUE7_BORPE	Q7vue7 bordetella	789	39	37.5	619	2	Q4CRE8_TRYCR	Q4cre8 trypanosoma
717	39	37.5	425	1	VIPA_SALTI	Q4972 salmonella	790	39	37.5	620	2	Q8U3Y3_PRRFU	Q8u3y3 pyrococcus
718	39	37.5	425	2	Q4IWZ8_AZOVI	Q4iww8 azotobacter	791	39	37.5	634	2	Q4UAS5_THEAN	Q4uas5 theleria a
719	39	37.5	425	2	Q9AP46_CITFR	Q9ap46 citrobacter	792	39	37.5	640	2	Q6MQ63_BDEBA	Q6mq63 bdellovibri
720	39	37.5	426	2	Q2LYX4_DROPS	Q2lyx4 drosophila	793	39	37.5	650	1	RRN3_PONPY	Q5r4n9 pongo pygma
721	39	37.5	426	2	Q4ATD9_9BURK	Q4atd9 polaronomas	794	39	37.5	651	1	RRN3_HUMAN	Q9nyv6 homo sapien
722	39	37.5	427	2	Q7RQZ2_PLAYO	Q7rtz2 plasmodium	795	39	37.5	652	2	Q4DI41_TRYCR	Q4di41 trypanosoma
723	39	37.5	428	2	Q61DG4_DROME	Q61dg4 drosophila	796	39	37.5	661	1	HSP7C_CAEBR	P19208 caenorhabdi
724	39	37.5	429	2	Q7VXU9_BORPE	Q7vxu9 bordetella	797	39	37.5	661	1	HSP7C_CAEBR	P27420 caenorhabdi
725	39	37.5	429	2	Q7W6F6_BORPA	Q7w6f6 bordetella	798	39	37.5	664	2	Q6CUV9_KLULA	Q6cuv9 kluyveromyc
726	39	37.5	429	2	Q7WIC6_BORBR	Q7wic6 bordetella	799	39	37.5	676	2	Q7S311_NEUCR	Q7s311 neurospora
727	39	37.5	431	2	Q3CPZ3_ALTAT	Q3cpz3 pseudolater	800	39	37.5	688	2	Q474X7_RALEJ	Q474x7 ralstonia e
728	39	37.5	433	1	Y036_SYNY3	Q55452 synchocyst	801	39	37.5	690	2	Q3S0L9_RALME	Q3s0l9 ralstonia m
729	39	37.5	433	2	Q26649_STRPU	Q26649 strongyloce	802	39	37.5	694	2	Q2XKH4_PSEPU	Q2xkh4 pseudomonas
730	39	37.5	433	2	Q2RTZ5_RHORU	Q2rtz5 rhodospirill	803	39	37.5	694	2	Q8BH50_PSEPK	Q8bh50 pseudomonas
731	39	37.5	437	2	Q2ZG09_CALSA	Q2zg09 caldicellul	804	39	37.5	695	2	Q4KD44_PSEF5	Q4kd44 pseudomonas
732	39	37.5	437	2	Q8EMC3_OCEIH	Q8emc3 oceanobacil	805	39	37.5	700	2	Q3FXA0_9BURK	Q3fxa0 rhodoferrax
733	39	37.5	439	1	HGD_DROME	Q8emc3 oceanobacil	806	39	37.5	706	2	Q9ADX5_9RH1Z	Q9adx5 agrobacteri
734	39	37.5	442	1	CHMO_AMATR	Q9xkj0 drosophila	807	39	37.5	747	2	Q59UP7_CANAL	Q59up7 candida alb
735	39	37.5	442	1	QNF1T1_9MICC	Q9jxel amaranthus	808	39	37.5	756	2	Q5IECT_ENTHI	Q5iect entamoeba h
736	39	37.5	443	2	Q8A2R2_BACTN	Q4nft1 athrobacte	809	39	37.5	762	1	E13B_TRIHA	P53626 trichoderma
737	39	37.5	453	2	Q5TTK6_ANOGA	Q8a2r2 bacteroides	810	39	37.5	762	2	Q8TG98_TRIVE	Q8tg98 trichoderma
738	39	37.5	455	2	Q5QLT6_ORYSA	Q5ttk6 anopheles g	811	39	37.5	765	2	Q3N755_9PROT	Q3n755 nitrosomona
739	39	37.5	457	2	Q6CON2_YARLI	Q5qlt6 oryza sativ	812	39	37.5	768	2	Q451W1_9BURK	Q451w1 burkholderi
740	39	37.5	457	2	Q7W1A5_BORPA	Q6con2 varrowia li	813	39	37.5	768	2	Q4LT17_9BURK	Q4lt17 burkholderi
741	39	37.5	457	2	MT36_OCEIH	Q7w1a5 bordetella	814	39	37.5	770	2	Q72U02_LEPIC	Q72u02 leptospira
742	39	37.5	460	1	Q8SUU0_ENCCU	Q7wp09 bordetella	815	39	37.5	770	2	Q8F152_LEPIN	Q8f152 leptospira
743	39	37.5	462	2	Q5FM17_LACAC	Q8e195 oceanobacil	816	39	37.5	773	2	Q452Z5_9BURK	Q452z5 burkholderi
744	39	37.5	463	2	Q5FMI7_LACAC	Q8suu0 encephaliti	817	39	37.5	773	2	Q4LNQ8_9BURK	Q4lnq8 burkholderi
745	39	37.5	464	2	Q4V4W8_DROME	Q5fmi7 lactobacill	818	39	37.5	777	2	Q3S9F3_BURS3	Q3s9f3 burkholderi
746	39	37.5	464	2	Q7XR50_ORYSA	Q4v4w8 drosophila	819	39	37.5	782	2	Q766X5_HUMAN	Q766x5 homo sapien
747	39	37.5	464	2	Q63BQ7_BACCH	Q81j40 plasmodium	820	39	37.5	783	2	Q4BDE2_BURVI	Q4bde2 burkholderi
748	39	37.5	474	2	Q6HJ76_BACHC	Q7xr50 oryza sativ	821	39	37.5	785	2	Q76D04_HUMAN	Q76d04 homo sapien
749	39	37.5	474	2	Q81QX6_BACAN	Q63bq7 bacillus ce	822	39	37.5	792	2	Q765H6_MOUSE	Q765h6 mus musculu
750	39	37.5	474	2	Q81QX6_BACAN	Q6h176 bacillus th	823	39	37.5	792	2	Q3V5L5_HUMAN	Q3v5l5 homo sapien
751	39	37.5	476	2	Q9ZSX1_MAIZE	Q81qx6 bacillus an	824	39	37.5	822	2	Q3PCZ4_PARDE	Q3pcz4 paracoccus
752	39	37.5	479	1	SCRB_STRMU	P13522 streptococc	825	39	37.5	829	2	Q3W6Y6_9ACTO	Q3w6y6 frankia sp.
753	39	37.5	481	1	SYC_GEOSL	Q7zext1 zea mays (m	826	39	37.5	831	2	Q8FXK3_BRUSU	Q8fxk3 bruceella su
754	39	37.5	483	2	Q7R3T9_GIALA	P13522 streptococc	827	39	37.5	831	2	Q8YDZ4_BRUME	Q8ydz4 bruceella me
755	39	37.5	484	2	Q3U9P0_HYDAT	Q747a2 geobacter s	828	39	37.5	831	2	Q2YIT8_BRUA2	Q2yit8 bruceella ab
756	39	37.5	486	2	Q38KD9_RHILU	Q9u9p0 hydra atten	829	39	37.5	831	2	Q9KIS9_BRUAB	Q9kis9 bruceella ab
757	39	37.5	487	2	Q44OG4_SOLUS	Q98kd9 rhizobium l	830	39	37.5	832	2	Q9RPI1_BRUSU	Q9rpi1 bruceella su
758	39	37.5	494	2	Q6ZZM4_STRAT	Q44og4 solibacter	831	39	37.5	844	2	Q44ZNS_SOLUS	Q44zn5 solibacter
759	39	37.5	504	2	Q8DIR1_SYNEL	Q6zzw4 streptomyce	832	39	37.5	846	2	Q3R071_XYLFA	Q3r071 xylella fas
760	39	37.5	505	2	Q69GG1_PPRAM	Q8dir1 synchococc	833	39	37.5	849	2	Q9PFP9_XYLFA	Q9pfp9 xylella fas
761	39	37.5	507	2	Q6Z4S9_ORYSA	Q69gg1 periplaneta	834	39	37.5				



835	39	37.5	858	2	Q8SW08_ENCCU	Q8sw08	encephalito	908	38	36.5	56	2	Q9DEL3_9TELE	Q9del3	coregonus s
836	39	37.5	897	2	Q38HX4_9CLOT	Q38hx4	clostridium	909	38	36.5	80	2	Q67LB3_SYMTH	Q67lb3	symbiobacte
837	39	37.5	898	2	Q9A749_CAUCR	Q9a749	caulobacter	910	38	36.5	91	2	Q3EB18_ACTSC	Q3ee18	actinobacil
838	39	37.5	916	2	Q8SIU2_ORYSA	Q8siu2	oryza sativ	911	38	36.5	92	2	Q8VWH5_SILDC	Q8vwh5	silene dicl
839	39	37.5	924	2	Q4HIP3_CAMLA	Q4hip3	campylobact	912	38	36.5	92	2	Q8W3Q8_SILDC	Q8w3q8	silene dicl
840	39	37.5	955	2	Q2RHH8_MOOTH	Q2rhh8	moorella th	913	38	36.5	92	2	Q8W3R4_SILLA	Q8w3r4	silene lati
841	39	37.5	980	2	Q5N8I9_ORYSA	Q5n8i9	oryza sativ	914	38	36.5	92	2	Q8M3U0_9CARY	Q8m3u0	silene ital
842	39	37.5	1013	2	Q337R2_ORYSA	Q337r2	oryza sativ	915	38	36.5	100	2	Q5ZU28_LEGPH	Q5zu28	legionella
843	39	37.5	1048	2	Q8YI36_MAIZE	Q8vy36	zea mays (m	916	38	36.5	109	2	Q3LOY7_9BACT	Q3loy7	uncultured
844	39	37.5	1068	2	Q24587_MAIZE	Q24587	zea mays (m	917	38	36.5	109	2	Q4E907_9RICK	Q4e907	wolbachia e
845	39	37.5	1087	2	Q6FXG2_CANGA	Q6fxg2	candida gla	918	38	36.5	113	2	Q2JXS3_9CVAN	Q2jxs3	cyanobacter
846	39	37.5	1154	2	Q5SC65_DICDI	Q5sc65	dictyosteli	919	38	36.5	118	2	Q4PCX1_USTMA	Q4pcx1	ustilago ma
847	39	37.5	1184	2	Q4DS96_TRYCR	Q4ds96	trypanosoma	920	38	36.5	121	2	Q9G5R0_9BILA	Q9g5r0	electra pil
848	39	37.5	1185	2	Q4D256_TRYCR	Q4d256	trypanosoma	921	38	36.5	128	2	Q80IW4_MEHVI	Q80iwa	meleagrid h
849	39	37.5	1185	2	Q8WJ93_LACPL	Q8wj93	lactobacill	922	38	36.5	134	1	VG32_ICHVI	Q00100	lctalurid h
850	39	37.5	1189	2	Q41D42_9BACI	Q41d42	exiguobacte	923	38	36.5	134	2	Q8Z174_YERPE	Q8z174	yersinia pe
851	39	37.5	1201	2	Q5BH33_EMENI	Q5bh33	aspergillus	924	38	36.5	135	2	Q37DQ3_RHOPA	Q37dq3	rhodospseud
852	39	37.5	1349	2	Q384J8_9TRYP	Q384j8	trypanosoma	925	38	36.5	143	2	Q3WID0_9ACTO	Q3wid0	frankia sp.
853	39	37.5	1374	2	Q621A7_CAEBR	Q621a7	caenorhabdi	926	38	36.5	143	2	Q4B9Q3_BURVI	Q4b9q3	burkholderi
854	39	37.5	1512	2	Q7XBD2_MAIZE	Q7xbd2	zea mays (m	927	38	36.5	143	2	Q2JC92_9ACTO	Q2jc92	frankia sp.
855	39	37.5	1740	2	Q2JBE1_9ACTO	Q2jeil	frankia sp.	928	38	36.5	148	2	Q03138_BAZTR	Q03138	bazzania tr
856	39	37.5	1941	2	Q44LS4_CHLLI	Q44ls4	chlorobium	929	38	36.5	150	2	Q6F7K3_ACTAD	Q6f7k3	acinatobact
857	39	37.5	2689	2	Q4ZRV1_PSEU2	Q4zrv1	pseudomonas	930	38	36.5	151	2	Q3IP69_NATPD	Q3ip69	natronomona
858	39	37.5	5432	2	Q5ZV52_9ACTO	Q5zv52	streptomyce	931	38	36.5	151	2	Q665U0_YERPS	Q665u0	yersinia ps
859	38.5	37.0	102	2	Q806X4_9HEPC	Q806x4	hepatitis c	932	38	36.5	153	2	Q8CKF9_YERPE	Q8ckf9	yersinia pe
860	38.5	37.0	177	2	Q58M30_9CAUD	Q58m30	cyanophage	933	38	36.5	159	2	Q9TT09_CERAB	Q9tt09	cercopithe
861	38.5	37.0	188	2	Q38EE8_9TRYP	Q38ee8	trypanosoma	934	38	36.5	162	2	Q6P655_HUMAN	Q6p655	homo sapien
862	38.5	37.0	188	2	Q450G4_9BURK	Q450g4	burkholderi	935	38	36.5	163	1	COAD_ENTFA	Q831P9	enteroococc
863	38.5	37.0	188	2	Q41V18_9BURK	Q41v18	burkholderi	936	38	36.5	168	1	ATSI_SCHPO	Q79081	schizosacch
864	38.5	37.0	224	2	Q6CLT9_YARLI	Q6clt9	yarrowia li	937	38	36.5	172	2	Q9QBL3_9TOMB	Q9qbl3	polaromonas
865	38.5	37.0	239	2	Q5DGV0_BRARE	Q5dgv0	brachydanio	938	38	36.5	172	2	Q9QBL3_9TOMB	Q9qbl3	tomato bush
866	38.5	37.0	252	2	Q4HLX9_CAMLA	Q4hlx9	campylobact	939	38	36.5	174	2	Q9HLJ3_9THEAC	Q9hlj3	thermoplasm
867	38.5	37.0	282	2	Q6ARS2_DESPS	Q6ar52	desulfotale	940	38	36.5	174	2	Q4X8H3_PLACH	Q4x8h3	stremotococ
868	38.5	37.0	288	2	Q4BR40_BURVI	Q4br40	burkholderi	941	38	36.5	183	2	Q6UZA9_STRPY	Q6uza9	bradyrhizob
869	38.5	37.0	344	2	Q3PHP1_PARDE	Q3php1	paracoccus	942	38	36.5	183	2	Q9AL12_ENTFA	Q9al12	entrococcu
870	38.5	37.0	358	2	Q5E1A4_VIBF1	Q5e1a4	vibrio fisc	943	38	36.5	189	2	Q30TB6_THIDN	Q30tb6	thiomicrosp
871	38.5	37.0	375	1	YL697_MIMIV	Q5unv8	mimivirus.	944	38	36.5	191	2	Q6ZL05_ORYSA	Q6zlo5	oryza sativ
872	38.5	37.0	387	2	Q4L6A3_STAHL	Q4l6a3	staphylococ	945	38	36.5	198	1	YNPB_YEAST	P53896	saccharomyc
873	38.5	37.0	387	2	Q6APX9_DESPS	Q6apx9	desulfotale	946	38	36.5	215	2	Q5FTQ4_GLUOX	P5ftq4	gluconobact
874	38.5	37.0	391	2	Q84W80_ARATH	Q84w80	arabidopsis	947	38	36.5	217	1	MSRA1_CAUCR	Q9a916	caulobacter
875	38.5	37.0	397	2	Q5JIS5_PYRKO	Q5jies5	pyrococcus	948	38	36.5	218	2	Q8W599_BRAJA	Q8w599	bradyrhizob
876	38.5	37.0	400	2	Q59022_PYRHO	Q59022	pyrococcus	949	38	36.5	219	2	Q5TQZ2_ANOGA	Q5tqz2	anopheles g
877	38.5	37.0	400	2	Q8U302_PYRFU	Q8u302	pyrococcus	950	38	36.5	229	2	Q4FP73_PELUB	Q4fp73	pelagibacte
878	38.5	37.0	400	2	Q9V0A7_PYRAB	Q9v0a7	pyrococcus	951	38	36.5	231	2	Q2IHL1_9DELT	Q2ihl1	anaeromyxob
879	38.5	37.0	453	2	Q68E17_BRARE	Q68e17	brachydanio	952	38	36.5	244	2	Q95ZL6_CABEL	Q95z16	caenorhabdi
880	38.5	37.0	478	2	Q3N6X7_9PROT	Q3n6x7	nitrosomona	953	38	36.5	244	2	Q378X1_RHOPA	Q378x1	rhodospseud
881	38.5	37.0	481	2	Q5SRP0_ARATH	Q5srp0	arabidopsis	954	38	36.5	245	2	Q344V6_RHOPA	Q344v6	rhodospseud
882	38.5	37.0	481	2	Q05619_PSESP	Q05619	pseudomonas	955	38	36.5	248	2	Q33F16_METHU	Q33f16	methanospir
883	38.5	37.0	482	2	Q48I29_PSE14	Q48i29	pseudomonas	956	38	36.5	252	2	Q95ZL7_CABEL	Q95z17	caenorhabdi
884	38.5	37.0	482	2	Q4IZ39_AZOVI	Q4iz39	azotobacter	957	38	36.5	260	2	Q3PE90_PARDE	Q3pe90	paracoccus
885	38.5	37.0	482	2	Q4ZSV6_PSEU2	Q4zsv6	pseudomonas	958	38	36.5	260	2	Q7VGY6_HELHP	Q7vgy6	helicobacte
886	38.5	37.0	482	2	Q881F0_PSESM	Q881f0	pseudomonas	959	38	36.5	260	2	Q3Y545_CYPCA	Q3y545	cyprinus ca
887	38.5	37.0	511	2	Q5SFB2_DICDI	Q5sfb2	dictyosteli	960	38	36.5	261	2	Q4XX23_PLACH	Q4xx23	plasmodium
888	38.5	37.0	526	2	Q69XEO_ORYSA	Q69xEO	oryza sativ	961	38	36.5	264	2	Q6DU96_XENTR	Q6dj96	xenopus tro
889	38.5	37.0	537	2	Q5WKY3_BACSK	Q5wky3	bacillus cl	962	38	36.5	265	2	Q3GG87_CHLVI	Q3gg87	prosthococh
890	38.5	37.0	539	2	Q99D91_9VIRU	Q99d91	disculla des	963	38	36.5	278	2	Q2W873_MAGSA	Q2w873	magnetospir
891	38.5	37.0	556	2	Q4W9D6_ASPFU	Q4w9d6	aspergillus	964	38	36.5	278	2	Q5HMA5_STAEO	Q5hma5	staphylococ
892	38.5	37.0	582	2	Q5J1R5_9NOCA	Q5j1r5	nocardia un	965	38	36.5	278	2	Q8CN18_STAES	Q8cn18	staphylococ
893	38.5	37.0	592	2	Q44DN1_CHRSL	Q44dn1	chromohalob	966	38	36.5	284	2	Q46L14_PROMT	Q46l14	prochloroco
894	38.5	37.0	597	1	NUOCD_BUCBP	Q895au4	buchnera ap	967	38	36.5	288	2	Q5K5C1_9VIRU	Q5k5c1	spodoptera
895	38.5	37.0	633	2	Q3GYK3_9ACTO	Q3gyk3	nocardioide	968	38	36.5	289	2	Q7Q5J5_ANOGA	Q7q5j5	anopheles g
896	38.5	37.0	665	2	Q4CTV3_TRYCR	Q4ctv3	trypanosoma	969	38	36.5	292	2	Q9N168_PAPHA	Q9n168	papio hamad
897	38.5	37.0	665	2	Q4DF54_TRYCR	Q4df54	trypanosoma	970	38	36.5	298	2	Q3IIU9_PSEHT	Q3iiu9	pseudodalter
898	38.5	37.0	695	2	Q74363_SCHPO	Q74363	schizosacch	971	38	36.5	302	2	Q886G1_PSESM	Q886g1	pseudomonas
899	38.5	37.0	732	2	Q8R6X3_THETN	Q8r6x3	thermoanaer	972	38	36.5	303	2	Q7NIB5_GLOVI	Q7nib5	gloeobacter
900	38.5	37.0	740	2	Q73J82_TREDE	Q73j82	treponema d	973	38	36.5	305	2	Q7NY51_CHRVO	Q7ny51	chromobacte
901	38.5	37.0	976	2	Q2NF92_9EURY	Q2nf92	methanospa	974	38	36.5	306	2	Q66GM9_ARATH	Q66gm9	arabidopsis
902	38.5	37.0	984	2	Q9LZV1_ARATH	Q9lztv1	arabidopsis	975	38	36.5	310	2	Q2RR18_RHORU	Q2rr18	rhodospiril
903	38.5	37.0	1457	2	Q3GEL2_9FIRM	Q3gel2	syntrophomo	976	38	36.5	313	1	Y1B2_HAEIN	P44554	haemophilus
904	38.5	37.0	1545	2	Q48866_DAUCA	Q48866	daucus caro	977	38	36.5	313	2	Q9SY10_ARATH	Q9sy10	arabidopsis
905	38.5	37.0	1761	2	Q48867_DAUCA	Q48867	daucus caro	978	38	36.5	316	2	Q4T6S6_TETNG	Q4t6s6	tetrarodon n
906	38	36.5	33	2	Q30547_PSEST	Q30547	pseudomonas	979	38	36.5	326	2	Q8XJ64_CLOPE	Q8xj64	clostridium
907	38	36.5	45	2	Q4E932_9RICK	Q4e932	wolbachia e	980	38	36.5	329	2	Q3FZQ7_9DELT	Q3fzq7	pelobacter

981 38 36.5 330 2 Q41UT3\_FERAC Q41ut3 ferroplasma  
 982 38 36.5 332 2 Q6LM33\_PROPR Q6lm33 photobacter  
 983 38 36.5 333 2 Q2P49\_RHIME Q2p49 rhizobium m  
 984 38 36.5 334 2 Q2J5W4\_FRANKA sp. Q2j5w4 frankia sp.  
 985 38 36.5 335 2 Q3NYA7\_9GAMM Q3nya7 shewanella  
 986 38 36.5 338 2 Q4R6K4\_MACFA Q4r6k4 macaca fasc  
 987 38 36.5 338 2 Q407C2\_JANNASCHIA Q407c2 jannaschia  
 988 38 36.5 338 2 Q9WZQ3\_THEMA Q9wzq3 thermotoga  
 989 38 36.5 341 1 P04970\_CAENORHABDI P04970 caenorhabdi  
 990 38 36.5 341 1 G3P4\_CAEEL P17331 caenorhabdi  
 991 38 36.5 341 2 Q4FSE8\_PSYAR Q4fse8 psychrobact  
 992 38 36.5 342 2 Q43Q84\_SOLUS Q43q84 solibacter  
 993 38 36.5 343 2 Q46823\_TRIASC Q46823 triakis scy  
 994 38 36.5 344 2 Q5FTK8\_GLUOX Q5ftk8 gluconobact  
 995 38 36.5 346 2 Q9SHG1\_ARATH Q9shg1 arabidopsis  
 996 38 36.5 349 2 Q3E8L3\_ARATH Q3e8l3 arabidopsis  
 997 38 36.5 350 2 Q6J678\_9BURK Q6j678 collimonas  
 998 38 36.5 353 2 Q08563\_RAT Q08563 rattus norv  
 999 38 36.5 354 2 Q4JV13\_CORJK Q4jv13 corynebacte  
 1000 38 36.5 355 2 Q37X36\_SPBAR Q37x36 novosphingo

## ALIGNMENTS

## RESULT 1

HPSE HUMAN STANDARD; PRT; 543 AA.  
 AC Q9Y251; Q53GB5; Q9UL39;  
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
 DT 01-NOV-1999, sequence version 1.  
 DT 07-FEB-2006, entry version 27.  
 DE Heparanase precursor (PC 3.2.-.-) (Heparanase-1) (Hpal) (Endo-  
 DE glucuronidase) (Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa  
 DE subunit).  
 GN Name=HPSE; Synonyms=HEP, HPA, HPA1, HPRI, HPSE1, HSE1;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.  
 RC TISSUE=Placenta;  
 RX MEDLINE=9335379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;  
 RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C.,  
 RA Seddon A.P., Giorgio N.A., Bohlen P.;  
 RT "Cloning and functional expression of a human heparanase gene.";  
 RL Biochem. Biophys. Res. Commun. 261:183-187(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN  
 RP SEQUENCE OF 158-168; 326-337 AND 447-491.  
 RC TISSUE=Embryonic fibroblast;  
 RX MEDLINE=93377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;  
 RA Toyoshima M., Nakajima M.;  
 RT "Human heparanase. Purification, characterization, cloning, and  
 RT expression.";  
 RL J. Biol. Chem. 274:24153-24160(1999).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.  
 RX PubMed=10395325; DOI=10.1038/10518;  
 RA Vlodavsky I., Friedmann Y., Elkin M., Aingorn H., Atzmon R.,  
 RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,  
 RA Spector L., Pecker I.;  
 RT "Mammalian heparanase: gene cloning, expression and function in tumor  
 RT progression and metastasis.";  
 RL Nat. Med. 5:793-802(1999).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE  
 RP OF 158-174; 263-272; 326-337; 433-436; 438-443; 466-468 AND 478-483.  
 RC TISSUE=Placenta;  
 RX MEDLINE=93321249; PubMed=10395326; DOI=10.1038/10525;  
 RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,

Parish C.R.;  
 RT "Cloning of mammalian heparanase, an important enzyme in tumor  
 RT invasion and metastasis.";  
 RL Nat. Med. 5:803-809(1999).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE  
 RP SPECIFICITY.  
 RC TISSUE=Placenta;  
 RX MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;  
 RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;  
 RT "Heparanase expression in invasive trophoblasts and acute vascular  
 RT damage.";  
 RL Glycobiology 10:467-475(2000).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.  
 RX PubMed=11547900; DOI=10.1023/A:1011375624902;  
 RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O.,  
 RA Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S.,  
 RA Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M.,  
 RA Farekh R., Stubberfield C.;  
 RT "Biochemical characterization of the active heterodimer form of human  
 RT heparanase (Hpal) protein expressed in insect cells.";  
 RL Biochem. J. 373:423-435(2003).  
 RN [8]  
 RP NUCLEOTIDE SEQUENCE [MRNA].  
 RA Pinnal M.A., Semedo P.;  
 RT "Cloned heparanase from MCF-7 cells.";  
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.  
 RN [9]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
 RC TISSUE=Small intestine;  
 RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,  
 RA Tanaka A., Yokoyama S.;  
 RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.  
 RN [10]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
 RC TISSUE=Pancreas;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywicki M.I., Skalska U., Smailus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [11]  
 RP MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.  
 RX PubMed=11123890; DOI=10.1021/bi002080p;  
 RA Hulett M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,  
 RA Gready J.E., Parish C.R.;

RT "Identification of active-site residues of the pro-metastatic  
RL endoglycosidase heparanase.";  
RL Biochemistry 39:15659-15667(2000).  
RN [12]  
RP N-GLYCOSYLATION, AND MUTAGENESIS OF ASN-162; ASN-178; ASN-200;  
RN ASN-217; ASN-238 AND ASN-459.  
RX PubMed=14573609; DOI=10.1074/jbc.M300541200;  
RA Simizu S., Iehida K., Wierzbicka M.K., Osada H.;  
RT "Secretion of heparanase protein is regulated by glycosylation in  
RL human tumor cell lines.";  
RL J. Biol. Chem. 279:2697-2703(2004).  
RN [13]  
RN SUBCELLULAR LOCATION.  
RP PubMed=15292202; DOI=10.1074/jbc.M402131200;  
RX Gingis-Velitski S., Zetser A., Kaplan V., Ben-Zaken O., Cohen E.,  
RA Levy-Adam F., Bashenko Y., Flugelman M.Y., Vlodavsky I., Ilan N.;  
RT "Heparanase uptake is mediated by cell membrane heparan sulfate  
RL proteoglycans.";  
RL J. Biol. Chem. 279:44084-44092(2004).  
RN [14]  
RN BIOPHYSICO-CHEMICAL PROPERTIES, PROCESSING, AND SUBCELLULAR LOCATION.  
RX PubMed=15948168; DOI=10.1016/j.febslet.2005.03.030;  
RA Cohen E., Atzmon R., Vlodavsky I., Ilan N.;  
RT "Heparanase processing by lysosomal/endosomal protein preparation.";  
RL FEBS Lett. 579:2334-2338(2005).  
RN [15]  
RN SUBCELLULAR LOCATION, PROCESSING, AND MUTAGENESIS OF TYR-156.  
RX PubMed=15659389; DOI=10.1074/jbc.M413370200;  
RA Aboud-Jarrous G., Rangini-Guetta Z., Aingorn H., Atzmon R.,  
RA Elgavish S., Peretz T., Vlodavsky I.;  
RT "Site-directed mutagenesis, proteolytic cleavage, and activation of  
RL human proheparanase.";  
RL J. Biol. Chem. 280:13568-13575(2005).  
RN [16]  
RN DOMAINS, AND MUTAGENESIS OF LYS-158 AND LYS-161.  
RX PubMed=15760902; DOI=10.1074/jbc.M414546200;  
RA Levy-Adam F., Aboud-Jarrous G., Guerrini M., Beccati D.,  
RA Vlodavsky I., Ilan N.;  
RT "Identification and characterization of heparin/heparan sulfate  
RL binding domains of the endoglycosidase heparanase.";  
RL J. Biol. Chem. 280:20457-20466(2005).  
RN [17]  
RN VARIANT SER-260.  
RX PubMed=15334672;  
RA Chen X.P., Liu Y.B., Rui J., Peng S.Y., Peng C.H., Zhou Z.Y.,  
RA Shi L.H., Shen H.W., Xu B.;  
RT "Heparanase mRNA expression and point mutation in hepatocellular  
RL carcinoma.";  
RL World J. Gastroenterol. 10:2795-2799(2004).  
CC -1- FUNCTION: Endoglycosidase which is a cell surface and  
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
CC proteoglycans. Also implicated in the extravasation of leukocytes  
CC and tumor cell lines. Due to its contribution to metastasis and  
CC angiogenesis, it is considered to be a potential target for anti-  
CC cancer therapies.  
CC -1- ENZYME REGULATION: Inhibited by EDTA, laminarin sulfate and, to a  
CC lower extent, by heparin and sulfamin and activated by calcium and  
CC magnesium (By similarity).  
CC -1- BIOPHYSICO-CHEMICAL PROPERTIES:  
CC pH dependence:  
CC Optimum pH is 4-6;  
CC -1- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
CC subunits, the proteolytic products.  
CC -1- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.  
CC Secreted, internalised and transferred to late endosomes/lysosomes  
CC as a proheparanase. In lysosomes, it is processed into the active  
CC form, the heparanase. The uptake or internalisation of  
CC proheparanase is mediated by HSPGs. Heparin appears to be a  
CC competitor and retain proheparanase in the extracellular medium.  
CC -1- TISSUE SPECIFICITY: Highly expressed in placenta and spleen and  
CC weakly expressed in lymph node, thymus, peripheral blood  
CC leukocytes, bone marrow, endothelial cells, fetal liver and tumor

CC tissues.  
CC -1- PTM: Proteolytically processed. The cleavage of the 65 kDa form  
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa  
CC product. The active form, the 8/50 kDa heterodimer, is resistant  
CC to degradation. Complete removal of the linker peptide appears to  
CC be a prerequisite to the complete activation of the enzyme.  
CC -1- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears  
CC to be essential for its solubility.  
  
Query Match 100.0%; Score 104; DB 1; Length 543;  
Best Local Similarity 100.0%; Pred. No. 9.7e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PAYLRFGGKTDFLIFDPK 19  
Db 89 PAYLRFGGKTDFLIFDPK 107  
|||||  
  
RESULT 2  
Q33X5 SPADJ  
ID Q33X5\_SPADJ PRELIMINARY; PRT; 558 AA.  
AC Q33X5;  
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.  
DT 06-DEC-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Heparanase.  
DE Name=hpa;  
OS Spalax judaei (Blind subterranean mole rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Spalacidae; Spalacinae; Spalax.  
OX NCBI\_TaxID=134510;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Aviavi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
RL cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Aviavi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
RL cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AM085494; CAJ30021.1; -; mRNA.  
SQ SEQUENCE 558 AA; 62737 MW; 07BAF8F55849EEE7 CRC64;  
  
Query Match 100.0%; Score 104; DB 2; Length 558;  
Best Local Similarity 100.0%; Pred. No. 1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PAYLRFGGKTDFLIFDPK 19  
Db 120 PAYLRFGGKTDFLIFDPK 138  
|||||  
  
RESULT 3  
Q33X6 SPADJ  
ID Q33X6\_SPADJ PRELIMINARY; PRT; 574 AA.  
AC Q33X6;  
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.  
DT 06-DEC-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Heparanase.  
DE Name=hpa;  
OS Spalax judaei (Blind subterranean mole rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Spalacidae; Spalacinae; Spalax.  
RN [1]  
RP NCBI_TaxID=134510;  
RC NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).  
CC -----  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AM085493; CAJ30020.1; -; mRNA.  
SQ SEQUENCE 574 AA; 64515 MW; 3AEBB13F07451684 CRC64;  
  
Query Match 100.0%; Score 104; DB 2; Length 574;  
Best Local Similarity 100.0%; Pred. No. 1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PAYLRFGGTGTDFLIFDPK 19  
Db 120 PAYLRFGGTGTDFLIFDPK 138  
|||||  
-----  
RESULT 4  
Q333X7_9RODE PRELIMINARY; PRT; 574 AA.  
ID Q333X7_9RODE PRELIMINARY; PRT; 574 AA.  
AC Q333X7;  
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.  
DT 06-DEC-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Heparanase.  
DE Name=hpa;  
OS Spalax carmeli.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Spalacidae; Spalacinae; Spalax.  
OX NCBI_TaxID=164324;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).  
CC -----  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AM085492; CAJ30019.1; -; mRNA.  
SQ SEQUENCE 574 AA; 64459 MW; 9F1D19DCBBD99DE CRC64;  
  
Query Match 100.0%; Score 104; DB 2; Length 574;  
Best Local Similarity 100.0%; Pred. No. 1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PAYLRFGGTGTDFLIFDPK 19  
Db 120 PAYLRFGGTGTDFLIFDPK 138  
|||||  
-----  
RESULT 5  
Q333X8_9RODE PRELIMINARY; PRT; 574 AA.  
ID Q333X8_9RODE PRELIMINARY; PRT; 574 AA.
```

```
AC Q333X8;  
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.  
DT 06-DEC-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Heparanase.  
DE Name=hpa;  
OS Spalax golani.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Spalacidae; Spalacinae; Spalax.  
OX NCBI_TaxID=191382;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).  
CC -----  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AM085491; CAJ30018.1; -; mRNA.  
SQ SEQUENCE 574 AA; 64555 MW; 48BEFE7D0BCB34 CRC64;  
  
Query Match 100.0%; Score 104; DB 2; Length 574;  
Best Local Similarity 100.0%; Pred. No. 1e-08;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 PAYLRFGGTGTDFLIFDPK 19  
Db 120 PAYLRFGGTGTDFLIFDPK 138  
|||||  
-----  
RESULT 6  
Q333X9_9RODE PRELIMINARY; PRT; 574 AA.  
ID Q333X9_9RODE PRELIMINARY; PRT; 574 AA.  
AC Q333X9;  
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.  
DT 06-DEC-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Heparanase.  
DE Name=hpa;  
OS Spalax galili.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Spalacidae; Spalacinae; Spalax.  
OX NCBI_TaxID=164323;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
cloning and identification of a novel splice variant.";  
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).  
CC -----  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----
```

DR EMBL; AM085490; CAJ30017.1; -; mRNA.  
 SQ SEQUENCE 574 AA; 64525 MW; 1635865051B380D0 CRC64;  
 Query Match 100.0%; Score 104; DB 2; Length 574;  
 Best Local Similarity 100.0%; Pred. NO. 1e-08;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PAYLRFGGTKDFLIFDPK 19  
 |||||  
 DB 120 PAYLRFGGTKDFLIFDPK 138  
 RESULT 7  
 HPSE MOUSE  
 ID HPSE MOUSE STANDARD; PRT; 535 AA.  
 AC Q6VGVZ1; Q8K3K3;  
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
 DT 11-OCT-2005, sequence version 2.  
 DT 07-MAR-2006, entry version 13.  
 DE Heparanase precursor (EC 3.2.2.-) (Endo-glucuronidase) [Contains:  
 DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].  
 DE Name=Hpspe; Synonyms=Hpa;  
 GN Name=Hpspe; Synonyms=Hpa;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA].  
 RC STRAIN=SJL/J; TISSUE=Spleen;  
 RX MEDLINE=93121249; PubMed=10395326; DOI=10.1038/10525;  
 RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,  
 RA Parish C.R.;  
 RT "Cloning of mammalian heparanase, an important enzyme in tumor  
 RT invasion and metastasis."  
 RL Nat. Med. 5:803-809(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 28-57 AND 150-179,  
 RP GLYCOSYLATION, BIOPHYSICOCHEMICAL PROPERTIES, ENZYME REGULATION, AND  
 RP SUBUNIT.  
 RC STRAIN=FVB; TISSUE=Embryo;  
 RX MEDLINE=22350326; PubMed=12460766; DOI=10.1016/S1046-5928(02)00558-2;  
 RA Miao H.-Q., Navarro E., Patel S., Sargent D., Koo H., Wan H.,  
 RA Plata A., Zhou Q., Ludwig T., Bohlen P., Kussie P.;  
 RT "Cloning, expression, and purification of mouse heparanase."  
 RL Protein Expr. Purif. 26:425-431(2002).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE [MRNA] AND ENZYME REGULATION.  
 RX MEDLINE=22841152; PubMed=12837765; DOI=10.1074/jbc.M300925200;  
 RA Gong F., Jemth P., Galvis M.L.E., Vlodavsky I., Horner A., Lindahl U.,  
 RA Li J.-P.;  
 RT "Processing of macromolecular heparin by heparanase."  
 RL J. Biol. Chem. 278:35152-35158(2003).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
 RC STRAIN=C57BL/6J, and NOD; TISSUE=Thymus;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
 RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 RA Di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA Georgi-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 RA Guetincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Hill D., Humnietek L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Keiso J., Kitamura H.,  
 RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,

Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 Matauda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,  
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
 RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 RA Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,  
 RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
 RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome."  
 RL Science 309:1559-1563(2005).  
 CC -!- FUNCTION: Endoglycosidase which is a cell surface and  
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
 CC proteoglycans. Also implicated in the extravasation of leukocytes  
 CC and tumor cell lines. Contributes to metastasis and angiogenesis  
 CC (By similarity).  
 CC -!- ENZYME REGULATION: Inhibited by EDTA and activated by calcium and  
 CC magnesium (By similarity). Inhibited by laminarin sulfate and, to  
 CC a lower extent, by heparin and sulfamin.  
 CC -!- BIOPHYSICOCHEMICAL PROPERTIES:  
 CC pH dependence:  
 CC Optimum pH is 5;  
 CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
 CC subunits, the proteolytic products.  
 CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.  
 CC Secreted, internalised and transferred to late endosomes/lysosomes  
 CC as a proheparanase. In lysosomes, it is processed into the active  
 CC form, the heparanase. The uptake or internalisation of  
 CC proheparanase is mediated by HSPGs. Heparin appears to be a  
 CC competitor and retain proheparanase in the extracellular medium  
 CC (By similarity).  
 CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form  
 CC leads to the generation of a linker peptide, 8 kDa and 50 kDa  
 CC product. The active form, the 8/50 kDa heterodimer, is resistant  
 CC to degradation. Complete removal of the linker peptide appears to  
 CC be a prerequisite to the complete activation of the enzyme (By  
 CC similarity).  
 CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears  
 CC to be essential for its solubility.  
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.  
 -----  
 Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 Distributed under the Creative Commons Attribution-NonDerivs License

EMBL; AF359507; AAQ15188.1; -; mRNA.  
 EMBL; AY077467; AAL76083.1; -; mRNA.  
 EMBL; AI151051; AAN41636.1; -; mRNA.  
 EMBL; AK040471; BAC30600.1; -; mRNA.  
 EMBL; AK154628; BAE32725.1; -; mRNA.  
 EMBL; ENSMUSG0000035273; Mus musculus.  
 MGI; MGI:1343124; Hpsae.  
 GO; GO:0005578; C,extracellular matrix (sensu Metazoa); TAS.  
 InterPro; IPR005199; Glyco\_hydro\_79\_N.  
 Pfam; PF03662; Glyco\_hydro\_79n; 1.  
 Calcium; Direct protein sequencing; Glycoprotein; Hydrolase; Lysosome;  
 Magnesium; Membrane; Signal.  
 SIGNAL 1 27 By similarity.  
 CHAIN 28 101 Heparanase 8 kDa subunit.  
 PROPEP 102 149 /FTRIDPRO.0000042263.  
 Linker peptide (By similarity).

```

FT CHAIN 150 535 /FTid=PRO_0000042264.
FT Heparanase 50 kDa subunit.
FT /FTid=PRO_0000042265.
FT Heparin/HS-binding (By similarity).
FT REGION 262 272 Heparin/HS-binding (By similarity).
FT REGION 217 217 Proton donor (Potential).
FT ACT_SITE 335 335 Nucleophile (Potential).
FT ACT_SITE 154 154 N-linked (GlcNAc. . .)
FT CARBOHYD 192 192 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 209 209 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 230 230 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 451 451 K -> R (in Ref. 3).
FT CONFLICT 206 206 W -> S (in Ref. 3).
FT CONFLICT 212 212 NGS -> DGL (in Ref. 1, 2 and 4).
FT CONFLICT 230 232 E -> K (in Ref. 3).
FT CONFLICT 335 335 G -> A (in Ref. 3).
FT CONFLICT 342 342 Y -> H (in Ref. 1, 2 and 4).
FT CONFLICT 455 455 V -> I (in Ref. 1, 2 and 4).
FT CONFLICT 531 531
FT SEQUENCE 535 AA; 60050 MW; AF19B28B7CD03F7B CRC64;

Query Match 95.2%; Score 99; DB 1; Length 535;
Best Local Similarity 100.0%; Pred. No. 6.9e-08; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0;

QY 1 PAYLRFGGTKTDFLI FDP 18
Db 81 PAYLRFGGTKTDFLI FDP 98

RESULT 8
HPSE RAT
ID HPSE RAT STANDARD; PRT; 536 AA.
AC Q1RPL; Q9QZF8;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 05-JUL-2004, sequence version 1.
DT 07-MAR-2006, entry version 11.
DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) [Contains:
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].
DE Name=Hpse; Synonyms=Hep;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
NP NUCLEOTIDE SEQUENCE [MRNA].
RP TISSUE=Placenta;
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;
RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
RA Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor
RT invasion and metastasis.";
RL Nat. Med. 5:803-809 (1999).
[2]
RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.
RX MEDLINE=22194309; PubMed=12077130; DOI=10.1074/jbc.M203282200;
RA Podyma-Inoue K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;
RT "Characterization of heparanase from a rat parathyroid cell line.";
J. Biol. Chem. 277:32459-32465 (2002).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Contributes to metastasis and angiogenesis
CC (By similarity).
CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower
CC extent, by heparin and sulfamin (By similarity). Activated by
CC calcium and magnesium. Inhibited by EDTA.
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products (By similarity).
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes

```

```

CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium
CC (By similarity).
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme (By
CC similarity).
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; AF359508; AAQ15189.1; -; mRNA.
DR EMBL; AF184967; AAF04563.1; -; mRNA.
DR RGD; 61969; Hpse.
DR InterPro; IPR005199; Glyco_hydro_79_N.
DR Pfam; PF03662; Glyco_hydro_79n; 1.
KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;
KW Signal.
FT SIGNAL 1 28 By similarity.
FT CHAIN 29 102 Heparanase 8 kDa subunit.
FT PROPEP 103 150 Linker peptide (By similarity).
FT CHAIN 151 536 /FTid=PRO_0000042267.
FT REGION 151 155 Heparanase 50 kDa subunit.
FT REGION 263 273 /FTid=PRO_0000042268.
FT ACT_SITE 218 218 Heparin/HS-binding (By similarity).
FT ACT_SITE 336 336 Proton donor (Potential).
FT CARBOHYD 155 155 Nucleophile (Potential).
FT CARBOHYD 193 193 N-linked (GlcNAc. . .) (By similarity).
FT CARBOHYD 210 210 N-linked (GlcNAc. . .) (By similarity).
FT CARBOHYD 452 452 N-linked (GlcNAc. . .) (By similarity).
FT CONFLICT 15 15 G -> R (in Ref. 2).
FT CONFLICT 227 227 H -> Q (in Ref. 2).
FT CONFLICT 350 350 D -> N (in Ref. 2).
SQ SEQUENCE 536 AA; 60480 MW; C434E04CF536EA4D CRC64;

Query Match 95.2%; Score 99; DB 1; Length 536;
Best Local Similarity 100.0%; Pred. No. 6.9e-08; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0;

QY 1 PAYLRFGGTKTDFLI FDP 18
Db 82 PAYLRFGGTKTDFLI FDP 99

RESULT 9
HPSE BOVIN
ID HPSE BOVIN STANDARD; PRT; 545 AA.
AC Q9MYI0;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-JUN-2001, sequence version 2.
DT 07-MAR-2006, entry version 15.
DE Heparanase precursor (EC 3.2.-.-) [Contains: Heparanase 8 kDa subunit;
DE Heparanase 50 kDa subunit].
DE Name=HPSE;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
[1]
RN NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RP TISSUE=Placenta;
RX MEDLINE=21176669; PubMed=11277877;

```







```
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jallion O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Desailly C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL: CAAE01012073; CAF94326.1; -: Genomic_DNA.
DR NON_TER 1
FT NON_TER 533
SQ SEQUENCE 533 AA; 60100 MW; 9B00A7C8780100FF CRC64;

Query Match 72.1%; Score 75; DB 2; Length 533;
Best Local Similarity 66.7%; Pred. No. 0.00092; Indels 0; Gaps 0;
Matches 12; Conservative 4; Mismatches 2;

QY 1 PAYLRFGGTKTDFLIFDP 18
||:|||||:|:|:|
Db 41 PAYLRFGGTRQDFNVFAP 58

RESULT 12
HPSE2_HUMAN STANDARD; PRT; 592 AA.
AC Q8HWQ2; Q5VUH4; Q5VUH5; Q5VUH6; Q8HWQ1; Q9HB37; Q9HB38; Q9HB39;
DT 25-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 25-OCT-2005, sequence version 2.
DT 07-MAR-2006, entry version 16.
DE Heparanase-2 (EC 3.2.-.-) (Hpa2).
GN Name=HPSE2; Synonyms=HPA2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA] (ISOFORMS 1; 3 AND 4), TISSUE SPECIFICITY,
RP AND SUBCELLULAR LOCATION.
RC TISSUE=Heart;
RX MEDLINE=20483645; PubMed=11027606; DOI=10.1006/bbr.2000.3586;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hircock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;
RT "Cloning and expression profiling of Hpa2, a novel mammalian
RT heparanase family member.";
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA] (ISOFORMS 1 AND 2).
RC TISSUE=Prostate;
RA Legoux P., Legoux R., O'Brien D., Salome M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA], AND VARIANT TYR-579.
RX PubMed=15164054; DOI=10.1038/nature02462;
```

```
RA Deloukas P., Earthworm M.E., Grafham D.V., Rubenfield M., French L.,
RA Steward C.A., Sims S.K., Jones M.C., Searle S., Scott C., Howe K.,
RA Hunt S.E., Andrews T.D., Gilbert J.G.R., Swarbreck D., Ashurst J.B.,
RA Taylor A., Battles J., Bird C.P., Ainscough R., Almeida J.P.,
RA Ashwell R.I.S., Ambrose K.D., Babbage A.K., Bagge C.L., Bailey J.,
RA Banerjee R., Bates K., Beasley H., Bray-Allen S., Brown A.J., Camire D.,
RA Brown J.Y., Burford D.C., Burrill W., Burton J., Cahill P., Clegg S.,
RA Carter N.P., Chapman J.C., Clark S.Y., Clarke G., Clee C.M., Clegg S.,
RA Corby N., Coulson A., Dhani P., Dutta I., Dunn M., Faulkner L.,
RA Frankish A., Frankland J.A., Garner P., Garnett J., Gribble S.,
RA Griffiths C., Grocock R., Gustafson E., Hammond S., Hawley J.L.,
RA Hart E., Heath P.D., Ho T.P., Hopkins B., Horne J., Howden P.J.,
RA Huckle E., Hynds C., Johnson C., Johnson D., Kana A., Kay M.,
RA Kimberley A.M., Kershaw J.K., Kokkinaki M., Laird G.K., Lawlor S.,
RA Lee H.M., Leongamornlert D.A., Laird G., Lloyd C., Lloyd D.M.,
RA Loveland J., Lovell J., McLaren S., McLay K.E., McMurray A.,
RA Mashreghi-Mohammadi M., Matthews L., Milne S., Nickerson T.,
RA Nguyen M., Overton-Larty E., Palmer S.A., Pearce A.V., Peck A.I.,
RA Pelan S., Phillimore B., Porter K., Rice C.M., Rogosin A., Ross M.T.,
RA Sarafidou T., Sehra H.K., Showkeen R., Skuce C.D., Smith M.,
RA Strandberg L., Sycamore N., Tester J., Thorpe A., Torcasso W.,
RA Tracey A., Tromans A., Tsolas J., Wall M., Walsh J., Wang H.,
RA Weinstock K., West A.P., Willey D.L., Whitehead S.L., Wilming L.,
RA Wray P.W., Young L., Chen Y., Lovering R.C., Moschonas N.K.,
RA Siebert R., Fehrel K., Bentley D., Durbin R., Hubbard T.,
RA Doucette-Stamm L., Beck S., Smith D.R., Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 10.";
RL Nature 429:375-381(2004).
CC -1- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Due to its contribution to metastasis and
CC angiogenesis, it is considered to be a potential target for anti-
CC cancer therapies.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1; Synonyms=HPA2c;
CC IsoId=Q8HWQ2-1; Sequences=Displayed;
CC Name=2;
CC IsoId=Q8HWQ2-2; Sequences=VSP_015852, VSP_015853;
CC Name=3; Synonyms=HPA2b;
CC IsoId=Q8HWQ2-3; Sequences=VSP_015851;
CC Name=4; Synonyms=HPA2a;
CC IsoId=Q8HWQ2-4; Sequences=VSP_015850;
CC -1- TISSUE SPECIFICITY: Widely expressed, with the highest expression
CC in brain, mammary gland, prostate, small intestine, testis and
CC uterus. Found both in normal and cancer tissues.
CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: AF282885; AAG23421.1; -: mRNA.
DR EMBL: AF282886; AAG23422.1; -: mRNA.
DR EMBL: AF282887; AAG23423.1; -: mRNA.
DR EMBL: AF299719; CAC82491.1; -: mRNA.
DR EMBL: AJ299720; CAC82492.1; -: mRNA.
DR EMBL: AL590036; CAH73137.1; -: Genomic DNA.
DR EMBL: AL139243; CAH73137.1; JOINED; Genomic DNA.
DR EMBL: AL356220; CAH73137.1; JOINED; Genomic DNA.
DR EMBL: AL356268; CAH73137.1; JOINED; Genomic DNA.
DR EMBL: AL445251; CAH73137.1; JOINED; Genomic DNA.
DR EMBL: AL139243; CAI14146.1; -: Genomic DNA.
DR EMBL: AL356220; CAI14146.1; JOINED; Genomic DNA.
DR EMBL: AL356268; CAI14146.1; JOINED; Genomic DNA.
DR EMBL: AL445251; CAI14146.1; JOINED; Genomic DNA.
DR EMBL: AL590036; CAI14146.1; JOINED; Genomic DNA.
DR EMBL: AL356268; CAH70448.1; -: Genomic DNA.
DR EMBL: AL139243; CAH70448.1; JOINED; Genomic DNA.
DR EMBL: AL356220; CAH70448.1; JOINED; Genomic DNA.
DR EMBL: AL445251; CAH70448.1; JOINED; Genomic DNA.
```

DR EMBL; AL590036; CAH70448.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAH16472.1; -; Genomic\_DNA.  
DR EMBL; AL139243; CAH16472.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356220; CAH16472.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAH16472.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAH16472.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356220; CAH17160.1; -; Genomic\_DNA.  
DR EMBL; AL139243; CAH17160.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAH17160.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAH17160.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAH17160.1; JOINED; Genomic\_DNA.  
DR EMBL; AL139243; CAH73139.1; -; Genomic\_DNA.  
DR EMBL; AL356220; CAH73139.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAH73139.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAH73139.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAH73139.1; JOINED; Genomic\_DNA.  
DR EMBL; AL139243; CAH70450.1; -; Genomic\_DNA.  
DR EMBL; AL356220; CAH70450.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAH70450.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAH70450.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAH70450.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAH16474.1; -; Genomic\_DNA.  
DR EMBL; AL139243; CAH16474.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356220; CAH16474.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAH16474.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAH16474.1; JOINED; Genomic\_DNA.  
DR EMBL; AL139243; CAH14148.1; -; Genomic\_DNA.  
DR EMBL; AL356220; CAH14148.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAH14148.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAH14148.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAH14148.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAH73138.1; -; Genomic\_DNA.  
DR EMBL; AL139243; CAH73138.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356220; CAH73138.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAH73138.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAH73138.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAH73138.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356220; CAH17161.1; -; Genomic\_DNA.  
DR EMBL; AL139243; CAH17161.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAH17161.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAH17161.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAH17161.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAH70449.1; -; Genomic\_DNA.  
DR EMBL; AL139243; CAH70449.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356220; CAH70449.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAH70449.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAH70449.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAH16473.1; -; Genomic\_DNA.  
DR EMBL; AL139243; CAH16473.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356220; CAH16473.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAH16473.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAH16473.1; JOINED; Genomic\_DNA.  
DR EMBL; AL139243; CAH14147.1; -; Genomic\_DNA.  
DR EMBL; AL356220; CAH14147.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAH14147.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAH14147.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAH14147.1; JOINED; Genomic\_DNA.  
DR PIR; JC7506; JC7506.  
DR Ensembl; ENSG00000172987; Homo sapiens.  
DR HGNC; HGNC:18374; HPSE2.  
DR GO; GO:0005622; C:intracellular; TAS.  
DR GO; GO:0030305; F:heparanase activity; TAS.  
DR InterPro; IPR005199; Glyco\_hydro\_79\_N.  
DR Pfam; PF03662; Glyco\_hydro\_79a; I.  
KW Alternative splicing; Hydrolase; Membrane; Polymorphism.  
CHAIN 1 592  
FT ACT SITE 262 262  
FT FT FT  
FT VARSPLIC 150 261  
FT Missing (in isoform 4).

FT VARSPLIC 204 261  
FT Missing (in isoform 3).  
FT FT  
FT VARSPLIC 539 548  
FT SVOLNQPLV -> TORCQVCGII (in isoform 2).  
FT FT  
FT VARSPLIC 549 592  
FT Missing (in isoform 2).  
FT FT  
FT VARSPLIC 579 579  
FT F -> Y (in dbSNP:10883100).  
FT FT  
FT CONFLICT 12 12  
FT P -> L (in Ref. 2; CAC82492).  
FT CONFLICT 213 213  
FT F -> S (in Ref. 2).  
SQ SEQUENCE 592 AA; 66580 MW; 95C384AD9A74258E CRC64;  
  
Query Match 63.5%; Score 66; DB 1; Length 592;  
Best Local Similarity 75.0%; Pred. No. 0.036; Mismatches 2; Indels 0; Gaps 0;  
Matches 12; Conservative 2;  
  
QY 1 PAYLRFGGTKDTPLIF 16  
||:|||||:|||||  
DB 110 PAFLRFGGRTDFLQF 125  
  
RESULT 13  
Q2M1H9\_HUMAN PRELIMINARY; PRT; 592 AA.  
AC Q2M1H9;  
DT 21-FEB-2006, integrated into UniProtKB/TrEMBL.  
DT 21-FEB-2006, sequence version 1.  
DT 21-FEB-2006, entry version 1.  
DE Heparanase 2.  
GN Name=HPSE2;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=PCR rescued clones;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong L.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smallos D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=PCR rescued clones;  
RG NIH MGC Project;  
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; BC112356; AAI12357.1; -; mRNA.  
SQ SEQUENCE 592 AA; 66610 MW; 94689E1C2A74359F CRC64;  
  
Query Match 63.5%; Score 66; DB 2; Length 592;

```

Best Local Similarity 75.0%; Pred. No. 0.036; Mismatches 2; Indels 0; Gaps 0;
Matches 12; Conservative 2;

QY 1 PAYLRFGGTKTDFLIF 16
DB 110 PAFLRFGGKRTDFLQF 125

RESULT 14
Q4TB80 TETNG
ID Q4TB80 TETNG PRELIMINARY; PRT; 597 AA.
AC Q4TB80; 2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Chromosome 17 SCAF7180, whole genome shotgun sequence. (Fragment).
ORFNames=GSFENG0003868001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontoidea; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
[1]
RN NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Blomont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
[2]
RN NUCLEOTIDE SEQUENCE.
RP Genoscope; Whitehead Institute Centre for Genome Research;
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; CAAB01007180; CAF89852.1; -; Genomic_DNA.
DR NON TER 597 597
FT SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;
SQ SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;

Query Match 63.5%; Score 66; DB 2; Length 597;
Best Local Similarity 75.0%; Pred. No. 0.037;
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLIF 16
DB 105 PAFLRFGGKRTDFLQF 120

RESULT 15
Q7MH05 VIBVY
ID Q7MH05 VIBVY PRELIMINARY; PRT; 326 AA.
AC Q7MH05;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 15-DEC-2003, sequence version 1.
DT 07-FEB-2006, entry version 17.
DE HflC protein.

```

```

GN OrderedLocusNames=VW3068;
OS Vibrio vulnificus (strain VJ016).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=196600;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=14656965; DOI=10.1101/gr.1295503;
RA Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Tsai H.-C.,
RA Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,
RA Shao C.-P., Lee C.-T., Hor L.-I., Tsai S.-F.;
RT "Comparative genome analysis of Vibrio vulnificus, a marine
RT pathogen.";
RL Genome Res. 13:2577-2587(2003).
-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; BA000037; BAC95832.1; -; Genomic_DNA.
DR Biocyc; VVUL196600:VV3068-MONOMER; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0005515; F:protein binding; IEA.
DR GO; GO:0008152; F:metabolism; IEA.
DR InterPro; IPR001544; Aminotrans_IV.
DR InterPro; IPR001107; Band 7.
DR InterPro; IPR10200; HflC_7.
DR PANTHER; PTHR10264; Band 7; 1.
DR Pfam; PF01145; Band_7; 1.
DR SMART; SM00244; PHB; 1.
DR TIGRFAMs; TIGR01932; hflC; 1.
KW Complete proteome.
SQ SEQUENCE 326 AA; 36883 MW; 2FC6D5F905CB22EB CRC64;

Query Match 51.0%; Score 53; DB 2; Length 326;
Best Local Similarity 55.6%; Pred. No. 3.2;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 AYLRFGGKTDPLIFDPK 19
DB 293 AYKSFSGTKDILVLDPK 310

RESULT 16
Q8DCU6 VIBVU
ID Q8DCU6 VIBVU PRELIMINARY; PRT; 326 AA.
AC Q8DCU6;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 14.
DE Membrane protease subunits.
GN OrderedLocusNames=VVL1297; ORFNames=VVL_1297;
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; AE016795; AA009752.1; -; Genomic_DNA.
DR Biocyc; VVUL216895:VVL1297-MONOMER; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.

```

```

DR GO: 0005515; P:protein binding; IEA.
DR GO: 0008152; P:metabolism; IEA.
DR InterPro: IPR001544; AminoTrans_IV.
DR InterPro: IPR001107; Band_7.
DR InterPro: IPR010200; HfIC.
DR PANTHER: PTHR10264; Band 7; 1.
DR Pfam: PF01145; Band 7; 1.
DR SMART: SM00244; PHB; 1.
DR TIGRFAMs: TIGR01932; HfIC; 1.
KW Complete proteome; Protease.
SQ SEQUENCE 326 AA; 36864 MW; 2FDA27F01BDEC2EB CRC64;

Query Match 51.0%; Score 53; DB 2; Length 326;
Best Local Similarity 55.6%; Pred. No. 3.2;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AYLRFGGKTDFLIFDPK 19
||:|||||:|
Db 293 AYKSFSGTKSDILVDPK 310

RESULT 17
Q72QP6 LEPIC
ID Q72QP6 LEPIC PRELIMINARY; PRT; 1538 AA.
AC Q72QP6;
DT 05-JUN-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUN-2004, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Hypothetical protein.
GN OrderedLocusNames=LIC12067; ORFNames=LIC_12067;
OS Leptospira interrogans serogroup Icterohaemorrhagiae serovar
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=44275;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Flocruz LI-130;
RX PubMed=15028702; DOI=10.1128/JB.186.7.2164-2172.2004;
RA Nascimento A.L.T.O., Ko A.I., Martins E.A.L., Monteiro-Vitorello C.B.,
RA Ho P.L., Haake D.A., Verjovski-Almeida S., Hartskeerl R.A.,
RA Marques M.V., Oliveira M.C., Meneck C.F.M., Leite L.C.C., Carrer H.,
RA Coutinho L.L., Degraeve W.M., Dellagostin O.A., El-Dorry H.,
RA Ferro E.S., Ferro M.I.T., Furlan L.R., Gamberini M., Gigliotti E.A.,
RA Goes-Neto A., Goldman G.H., Goldman M.H.S., Harakava R.,
RA Jeronimo S.M.B., Junqueira-de-Azevedo I.L.M., Kimura E.T.,
RA Kuramae E.E., Lemos E.G.M., Lemos M.V.F., Marino C.L., Nunes L.R.,
RA de Oliveira R.C., Pereira G.G., Reis M.S., Schrieffer A.,
RA Siqueira W.J., Sommer P., Tsai S.M., Simpson A.J.G., Ferro J.A.,
RA Camargo L.E.A., Kitajima J.P., Satubal J.C., Van Sluys M.A.;
RT "Comparative genomics of two Leptospira interrogans serovars reveals
RT novel insights into physiology and pathogenesis.";
RL J. Bacteriol. 186:2164-2172(2004).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License

DR EMBL: A616823; AAS70638.1; -: Genomic DNA.
DR InterPro: IPR001589; Actnin actin bd.
DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 1538 AA; 183383 MW; 6FF9B985E45A821C CRC64;

Query Match 51.0%; Score 53; DB 2; Length 1538;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFLIFDP 18
||:|||||:|
Db 486 PFYFRFGDDLDDYVSFSP 503

RESULT 18
Q8F5E5_LEPIN
ID Q8F5E5_LEPIN PRELIMINARY; PRT; 1538 AA.
AC Q8F5E5;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-MAR-2006, entry version 14.
DE Hypothetical protein.
GN OrderedLocusNames=LAL1737; ORFNames=LA_1737;
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_TaxID=173;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RX MEDLINE=22598143; PubMed=12712204; DOI=10.1038/nature01597;
RA Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,
RA Zhang Y.-X., Xiong H., Lu G., Lu L.-F., Jiang H.-Q., Jia J., Tu Y.-F.,
RA Jiang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,
RA Zhang Y., Zhu G.-F., Wan M., Huang H.-L., Qian Z., Wang S.-Y., Ma W.,
RA Yao Z.-J., Shen Y., Qiang B.-Q., Xia Q.-C., Guo X.-K., Danchin A.,
RA Saint Girons I., Somerville R.L., Wen Y.-W., Shi M.-H., Chen Z.,
RA Xu J.-G., Zhao G.-P.;
RT "Unique physiological and pathogenic features of Leptospira
RT interrogans revealed by whole-genome sequencing.";
RL Nature 422:888-893(2003).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License

DR EMBL: AE010300; AAN48936.1; -: Genomic DNA.
DR GenomeReviews: AE010300.GR; LAL1737.
DR BioCyc: LINT189518:LAL1737-MONOMER; -.
DR InterPro: IPR001589; Actnin actin bd.
DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 1538 AA; 183389 MW; 5E3A7868F815FA0F CRC64;

Query Match 51.0%; Score 53; DB 2; Length 1538;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFLIFDP 18
||:|||||:|
Db 486 PFYFRFGDDLDDYVSFSP 503

RESULT 19
Q8LHX2_ORYSA
ID Q8LHX2_ORYSA PRELIMINARY; PRT; 512 AA.
AC Q8LHX2;
DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2002, sequence version 1.
DT 07-FEB-2006, entry version 9.
DE Hypothetical protein P0022B05.130 (Hypothetical protein
DE OSUNBA0057M23.102).
GN Name=P0022B05.130; Synonyms=OSJNBA0057M23.102;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;
OC Ehrhartoideae; Oryzoideae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
RT clone:OSJNBA0057M23.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License

```

```
CC -----
DR EMBL; AP004262; BAC10824.1; -; Genomic_DNA.
DR EMBL; AP005165; BAD30945.1; -; Genomic_DNA.
DR Gramene; Q8LHX2; -.
DR InterPro; IPR013181; U.
KW Hypothetical protein.
SQ SEQUENCE 512 AA; 58686 MW; 6E4C74892192BD43 CRC64;

Query Match 48.1%; Score 50; DB 2; Length 512;
Best Local Similarity 50.0%; Pred. No. 17;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 YLRFGGKTDFLIIFDP 18
   : ||||| : |||
Db 146 FVEFGGTPHQVLFDDP 161

RESULT 20
Q4HLX4 CAMLA
ID Q4HLX4_CAMLA PRELIMINARY; PRT; 224 AA.
AC Q4HLX4;
DT 16-AUG-2005, integrated into UniProtKB/TrEMBL.
DT 16-AUG-2005, sequence version 1.
DT 21-FEB-2006, entry version 6.
DE Response regulator.
GN ORFNames=CLA0531;
OS Campylobacter lari RM2100.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=306263;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=RM2100;
RA Fouts D.E.; Mongodin E.F.; Mandrell R.E.; Miller W.G.; Rasko D.A.;
RA Jacques R.J.; Brinkac L.M.; DeBoy R.T.; Parker C.T.; Daugherty S.C.;
RA Dodson R.J.; Durkin A.S.; Madupu R.R.; Sullivan S.A.; Shetty J.U.;
RA Ayodeji M.A.; Shvartsbeyn A.A.; Schatz M.C.; Badger J.H.; Fraser C.M.;
RA Nelson K.E.;
RT "Major structural and novel potential virulence mechanisms from the
RT genomes of multiple Campylobacter species.";
RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -1- SUBCELLULAR LOCATION: Cytoplasm (By similarity).
CC -1- SIMILARITY: Contains 1 response regulatory domain.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAFK01000002; EAL55101.1; -; Genomic_DNA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0000156; F:two-component response regulator activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0007600; P:sensory perception; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0000160; P:two-component signal transduction system (p. .; IEA.
DR InterPro; IPR001789; Response_reg.
DR InterPro; IPR001867; Trans_reg_C.
DR Pfam; PF00072; Response_reg_1.
DR Pfam; PF00486; Trans_reg_C; 1.
DR ProDom; PD0000039; Response_reg; 1.
DR ProDom; PD0000329; Trans_reg_C; 1.
DR SMART; SM00448; REC; 1.
DR PROSITE; PS01110; RESPONSE_REGULATORY; 1.
DR KW Activator; DNA-binding; Phosphorylation; Sensory transduction;
KW Transcription; Transcription regulation;
KW Two-component regulatory system.
SQ SEQUENCE 224 AA; 25674 MW; 0CD80C28FE6D152 CRC64;

Query Match 47.6%; Score 49.5; DB 2; Length 224;
Best Local Similarity 60.0%; Pred. No. 8.4;
Matches 12; Conservative 1; Mismatches 4; Indels 3; Gaps 1;

QY 2 AYLRFGGT---KTDFLIIFDP 18
   : ||||| : |||
Db 114 ARLRFGGTNNIKIDLDVIDP 133

RESULT 21
Q6Z4S8 ORYSA
ID Q6Z4S8_ORYSA PRELIMINARY; PRT; 513 AA.
AC Q6Z4S8;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 7.
DE Hypothetical protein OSJNBa0057M23.122;
DE Name=OSJNBa0057M23.122;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;
OC Ehrhartoideae; Oryzoae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T.; Matsumoto T.; Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
RT clone:OSJNBa0057M23.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AP005165; BAC83785.1; -; Genomic_DNA.
DR Gramene; Q6Z4S8; -.
DR InterPro; IPR013181; U.
KW Hypothetical protein.
SQ SEQUENCE 513 AA; 59130 MW; D94C59E999A036EFD CRC64;

Query Match 47.1%; Score 49; DB 2; Length 513;
Best Local Similarity 60.0%; Pred. No. 26;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 4 LRFGGKTDFLIIFDP 18
   : ||||| : |||
Db 147 LEFGGTPRRYLFDDP 161

RESULT 22
O67183 AQUAE
ID O67183_AQUAE PRELIMINARY; PRT; 588 AA.
AC O67183;
DT 01-AUG-1998, integrated into UniProtKB/TrEMBL.
DT 01-AUG-1998, sequence version 1.
DT 07-FEB-2006, entry version 22.
DE Mannosyltransferase A.
GN Name=mtfA; OrderedLocustNames=AQ_1096;
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=VFS;
RX MEDLINE=98196666; PubMed=9537320; DOI=10.1038/32831;
RA Deckert G.; Warren P.V.; Gaasterland T.; Young W.G.; Lenox A.L.;
RA Graham D.E.; Overbeek R.; Sneed M.A.; Keller M.; Aujay M.; Huber R.;
RA Feldman R.A.; Short J.M.; Olsen G.J.; Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AE000657; AAC07142.1; -; Genomic_DNA.
DR PIR; D70394; D70394.
```

```
DR BIOCYC: RAE063363:AO_1096-MONOMER; -.
DR GO: GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR CO: CO:0009058; P:biosynthesis; IEA.
DR InterPro: IPR001296; Glyco_transf_1.
DR Pfam: PF00534; Glycosyltransferase; Transferase.
KW Complete proteome; Glycosyltransferase; Transferase.
SQ SEQUENCE 588 AA; 69408 MW; E47365CA903894EF CRC64;

Query Match 47.1%; Score 49; DB 2; Length 588;
Best Local Similarity 60.0%; Pred. No. 30;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 YLRFGGTKTDFLIFD 17
||:| ||| ||:|
Db 307 YLRFKKTKTDFLIFD 321

RESULT 23
Q8PS30 METWA PRELIMINARY; PRT; 161 AA.
AC Q8PS30;
DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2002, sequence version 1.
DT 07-FEB-2006, entry version 11.
DE Conserved protein.
GN ORFNames=MA_3257;
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Goel / GoI / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Depkenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wierzer A., Baeumer S., Jacobi C.,
RA Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazei: evidence for lateral gene
transfer between Bacteria and Archaea.";
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DE Hypothetical protein.
GN ORFNames=MA_3400;
EMBL: AE008384; AA032953.1; -: Genomic_DNA.
DR BIOCYC: MAA2192952:MM3257-MONOMER; -.
KW Complete proteome.
SQ SEQUENCE 161 AA; 18820 MW; 23846110E3204F0B CRC64;

Query Match 46.2%; Score 48; DB 2; Length 161;
Best Local Similarity 52.9%; Pred. No. 11;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFLIFD 17
||| ||| ||:|
Db 97 PKDARFLGTPVDFIVFD 113

RESULT 24
Q8TKK5 METAC PRELIMINARY; PRT; 176 AA.
AC Q8TKK5;
DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2002, sequence version 1.
DT 07-MAR-2006, entry version 11.
DE Hypothetical protein.
GN ORFNames=MA_3400;
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;

DR BIOCYC: RAE063363:AO_1096-MONOMER; -.
DR STRAIN=C2A / ATCC 35395 / DSM 2834;
DR MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettnier H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
and physiological diversity.";
RL Genome Res. 12:532-542(2002).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DE Hypothetical protein.
GN ORFNames=MA_3257;
EMBL: AE010299; AA06767.1; -: Genomic_DNA.
DR BIOCYC: MACE18937:MA3400-MONOMER; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 176 AA; 20594 MW; 19576D9B075E62C8 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 176;
Best Local Similarity 52.9%; Pred. No. 12;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFLIFD 17
||| ||| ||:|
Db 112 PKDARFLGTPVDFIVFD 128

RESULT 25
Q2JIZ0 9CYAN PRELIMINARY; PRT; 194 AA.
ID Q2JIZ0;
AC Q2JIZ0;
DT 07-MAR-2006, integrated into UniProtKB/TrEMBL.
DT 07-MAR-2006, sequence version 1.
DT 07-MAR-2006, entry version 1.
DE Hypothetical protein.
GN ORFNames=CVB_2470;
OS Cyanobacteria bacterium Yellowstone B-Prime.
OC Bacteria; Cyanobacteria.
OX NCBI_TaxID=331115;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Yellowstone B-Prime;
RA Heidelberg J.;
RL Submitted (DEC-2005) to the EMBL/GenBank/DBDJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DE Hypothetical protein.
GN ORFNames=CP000240;
EMBL: CP000240; AB03404.1; -: Genomic_DNA.
DR BIOCYC: CP000240; AB03404.1; -: Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 194 AA; 22088 MW; 0DA56BDEA16427D CRC64;

Query Match 46.2%; Score 48; DB 2; Length 194;
Best Local Similarity 58.8%; Pred. No. 13;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 PAYLRFGGTKTDFLIFD 17
||| ||| ||:|
Db 117 POOLAFQKQGDFLIFE 133

RESULT 26
Q86YUQ_HUMAN
```

```
ID Q86YU0_HUMAN PRELIMINARY; PRT; 208 AA.
AC Q86YU0;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Carbonic anhydrase VII short form.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RP Chen Y., Huang C.-H.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AY075020; AAL78168.1; -; mRNA.
DR HSP; P00918; 1UGD.
DR Ensembl; ENSG00000168748; Homo sapiens.
DR GO; GO:0004089; F:carbonate dehydratase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006730; P:one-carbon compound metabolism; IEA.
DR InterPro; IPR001148; Euk_Coanhd.
DR PANTHER; PTHR18952; Euk_Coanhd.
DR Pfam; PF00194; Carb anhydrase; 1.
DR ProDom; PD000865; Euk_Coanhd; 1.
DR PROSITE; PS00162; ALPHA CA 1; 1.
DR SQ SEQUENCE 208 AA; 23452 MW; AF3D016A27182D18 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 208;
Best Local Similarity 56.2%; Pred. No. 14;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 LRFGGTKTDFLIIDPK 19
   :||| ||| |.|||
Db 113 VRFKGTAKQSFNPK 128

RESULT 27
ID Q2RP20_RHURU PRELIMINARY; PRT; 247 AA.
AC Q2RP20;
DT 24-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 24-JAN-2006, sequence version 1.
DT 07-FEB-2006, entry version 2.
DE Cytochrome-C oxidase, mono-heme subunit/FixO.
GN ORFNames=Rru_A3331;
OS Rhodospirillum rubrum ATCC 11170.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Rhodospirillaceae; Rhodospirillum.
OX NCBI_TaxID=269796;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 11170;
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Munk A.C., Brettin T., Bruce D.,
RA Han C., Tapia R., Gilna P., Schmutz J., Larimer F., Land M.,
RA Kyrpides N., Mavromatis K., Richardson P., Zhang Y., Roberts G.,
RA Reslewic S., Zhou S., Schwartz D.C.;
RT "Complete sequence of the chromosome of Rhodospirillum rubrum ATCC
RT 11170."
RL Submitted (DEC-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; CP000230; ABC24125.1; -; Genomic DNA.
DR SQ SEQUENCE 247 AA; 27561 MW; 359BFD038B0483A1 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 247;
```

```
Best Local Similarity 58.8%; Pred. No. 17;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 AYLRFGGKTDFLIIDP 18
   |||: ||| |||
Db 224 AYQLVLGTMTVDFTFDP 240

RESULT 28
ID CAH7_HUMAN STANDARD; PRT; 264 AA.
AC P43166; Q541F0;
DT 01-NOV-1995, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1995, sequence version 1.
DT 07-FEB-2006, entry version 44.
DE Carbonic anhydrase 7 (EC 4.2.1.1) (Carbonic anhydrase VII) (Carbonate
DE dehydratase VII) (CA-VII).
GN Name=CA7;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RX MEDLINE=92147127; PubMed=1783392;
RA Montgomery J.C., Venta P.J., Eddy R.L., Fukushima Y.S., Shows T.B.,
RA Tashian R.E.;
RT "Characterization of the human gene for a newly discovered carbonic
RT anhydrase, CA VII, and its localization to chromosome 16."
RL Genomics 11:835-848(1991).
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA].
RP Chen Y., Huang C.-H.;
RT "Molecular identification of carbonic anhydrases (CA) and CA-related
RT (CAR) genes."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX MEDLINE=99425270; PubMed=10493829; DOI=10.1006/geno.1999.5927;
RA Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
RA Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
RA Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT "Genome duplications and other features in 12 Mb of DNA sequence from
RT human chromosome 16p and 16q."
RL Genomics 60:295-308(1999).
RN [4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Colon, kidney, and stomach;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ussdin T.B., Ioshizuka S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grumwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Reversible hydration of carbon dioxide.
CC -!- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
```



```
CC -|- COPACTOR: Zinc (By similarity).
CC -|- SUBCELLULAR LOCATION: Cytoplasm (Probable).
CC -|- SIMILARITY: Belongs to the alpha-carbonic anhydrase family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC -----
CC EMBL: M76423; AAAS1923.1; -; Genomic DNA.
CC EMBL: M76420; AAAS1923.1; JOINED; Genomic DNA.
CC EMBL: M76421; AAAS1923.1; JOINED; Genomic DNA.
CC EMBL: M76422; AAAS1923.1; JOINED; Genomic DNA.
CC EMBL: AY075019; AAL78167.1; -; mRNA.
CC EMBL: AC004638; AAC23785.1; -; Genomic DNA.
CC EMBL: BC033865; AAH33865.1; -; mRNA.
CC PIR: A55272; CRU07.
CC HSPF; P00918; IBV3.
CC Ensembl; ENSG00000168748; Homo sapiens.
CC HGNC; HGNC:1381; CAV.
CC MIM; 114770; Gene.
CC GO; GO:0004089; F:carbonate dehydratase activity; TAS.
CC InterPro; IPR001148; Euk Coanhd.
CC PANTHER; PTHR18952; Euk Coanhd; 1.
CC Pfam; PF00194; Carb. anhydrase; 1.
CC ProDom; PD000865; Euk Coanhd; 1.
CC PROSITE; PS00162; ALPHA_CA_1; 1.
CC PROSITE; PS51144; ALPHA_CA_2; 1.
CC Lyase; Metal-binding; Zinc.
CC CHAIN 1 264 Carbonic anhydrase 7.
FT METAL 96 96 /FTID=PRO_0000077431.
FT METAL 98 98 Zinc (catalytic) (By similarity).
FT METAL 121 121 Zinc (catalytic) (By similarity).
SQ SEQUENCE 284 AA; 29658 MW; 7AD559FC6E07EF96 CRC64;

Query Match 46.2%; Score 48; DB 1; Length 264;
Best Local Similarity 56.2%; Pred. No. 18;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 LRFGGKTKDFLIIDPK 19
Db 169 VRFKGTRAQSCFNPK 184

RESULT 29
Q928Y1 CHLPN PRELIMINARY; PRT; 265 AA.
AC Q928Y1; 07AJ65; Q7BXP9; Q7DEH8;
DT 01-MAY-1999, integrated into UniProtKB/TrEMBL.
DT 01-MAY-1999, sequence version 1.
DT 21-FEB-2006, entry version 30.
DE Hypothetical protein CP0564 (Hypothetical protein CPJ0203).
GN OrderedLocustNames=CP0564, CPJ0203, CPB0207; ORFNames=CPn_0203;
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=CML029;
RX MEDLINE=99206606; PubMed=10192388; DOI=10.1038/7716;
RA Kaiman S., Mitchell W.P., Marathe R., Lammel C.J., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389 (1999).
RN [2]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935; DOI=10.1093/nar/28.6.1397;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J.D., Uterback T.R., Berry K.J.,
RA Baas S., Linher K.D., Weidman J.F., Khouri H.M., Craven B., Bowman C.,
RA Dodson R.J., Gwin M.L., Nelson W.C., DeBoy R.T., Kolonay J.F.,
RA McClarty G., Salzberg S.L., Eisen J.A., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
```

```
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406 (2000).
RN [3]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=TW-183;
RA Geng M.M., Schuhmacher A., Muehlendorfer I., Bensch K.W., Schaefer K.P.,
RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;
RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with
RT other Chlamydia strains based on whole genome sequence analysis.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362; DOI=10.1093/nar/28.12.2311;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CML029 from USA.";
RL Nucleic Acids Res. 28:2311-2314 (2000).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC -----
CC EMBL: AE001363; AAD18356.1; -; Genomic DNA.
CC EMBL: AE002215; AAP38383.1; -; Genomic DNA.
CC EMBL: AE017157; AAP98140.1; -; Genomic DNA.
CC EMBL: BA000008; BAA98413.1; -; Genomic DNA.
CC PIR; C86516; C86516.
CC PIR; D72105; D72105.
CC TIGR; CP0564; -.
CC BioCyc; CPNE115711:CP0564-MONOME-; -.
CC BioCyc; CPNE115713:CPN0203-MONOM-; -.
CC BioCyc; CPNE138677:CPN0203-MONOM-; -.
CC BioCyc; CPNE182082:CPN0207-MONOM-; -.
CC InterPro; IPR006974; DUF648.
CC Pfam; PF04890; DUF648; 1.
CC Complete proteome; Hypothetical protein.
SQ SEQUENCE 265 AA; 30482 MW; 36B4C6B4055267A1 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 265;
Best Local Similarity 47.1%; Pred. No. 18;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 AYLRFGGKTKDFLIIDFP 18
Db 24 SYFFFGGTRTQILVITP 40

RESULT 30
Q41BD7_9BACI PRELIMINARY; PRT; 311 AA.
AC Q41BD7;
DT 27-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 27-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE PfkB.
GN ORFNames=ExigDRAFT_0369;
OS Exiguobacterium sp. 255-15.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Exiguobacterium.
OX NCBI_TaxID=262543;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=255-15;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Izrani S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Exiguobacterium sp.
RT 255-15.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=255-15;
RG US DOE Joint Genome Institute (JGI-ORNL);
```

```
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Exiguobacterium sp. 255-15";
RN Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RL [3]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=255-15;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T., Hammon N., Israni S., Pitluck S., Richardson P.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AADW02000018; EAM86641.1; -; Genomic_DNA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0004747; F:ribokinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006014; P:D-ribose metabolism; IEA.
DR InterPro; IPR002173; PfKB.
DR InterPro; IPR011611; PfKB region.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; PfKB; 1.
DR PRINTS; PR00990; RIBOKINASE.
DR PROSITE; PS00583; PFKB_KINASES_1; 1.
DR PROSITE; PS00584; PFKB_KINASES_2; 1.
KW Kinase; Transferase.
SQ SEQUENCE 311 AA; 33430 MW; 37F2CA05DD010579 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 311;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 8 GTKTDLFLFDPK 19
DB 68 GVKTDHLVDFPE 79

RESULT 31
ID Q8G431_BIFLO PRELIMINARY; PRT; 411 AA.
AC Q8G431;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 16.
DE Probable aminotransferase.
GN OrderedLocustNames=BL1564;
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787; DOI=10.1073/pnas.212527599;
RA Schell M.A., Karimantou M., Snel B., Vilanova D., Berger B., Pessi G., Zahren M.-C., Desiere F., Bork P., Delley M., Fridmore R.D., Arigoni F.;
RL "The genome sequence of Bifidobacterium longum reflects its adaptation to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427 (2002).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AE014295; AAN25355.1; -; Genomic_DNA.
DR BioCyc; BLON206672:BL1564-MONOMER; -.
DR GO; GO:0008483; F:transaminase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
```

```
DR InterPro; IPR004839; Aminotrans_I/II.
KW Aminotransferase; Complete proteome; Transferase.
SQ SEQUENCE 411 AA; 43070 MW; EC2F7E468523A26D CRC64;

Query Match 46.2%; Score 48; DB 2; Length 411;
Best Local Similarity 75.0%; Pred. No. 30;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAYLRFPGTKTD 12
DB 388 PAYLRFSAFATD 399

RESULT 32
ID Q3FBC5_9BURK PRELIMINARY; PRT; 425 AA.
AC Q3FBC5;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Mandelate racemase/muconate lactonizing enzyme.
GN ORFNames=BambDRAFT_3416;
OS Burkholderia ambifaria AMMD.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; Burkholderia cepacia complex.
OX NCBI_TaxID=339670;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AMMD;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T., Hammon N., Israni S., Pitluck S., Richardson P.;
RL "Sequencing of the draft genome and assembly of Burkholderia ambifaria AMMD.";
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AMMD;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RL "Annotation of the draft genome assembly of Burkholderia ambifaria AMMD.";
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAJ01000005; EAO46342.1; -; Genomic_DNA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001354; MR_MLE.
DR Pfam; PF01188; MR_MLE; 1.
SQ SEQUENCE 425 AA; 45842 MW; CC9E0BB8836F7C32 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 425;
Best Local Similarity 56.2%; Pred. No. 31;
Matches 9; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 5 RFGGTTK--DFLIFDP 18
DB 313 RYGLLRADRDFLVDFP 328

RESULT 33
ID Q8T108_BOMMO PRELIMINARY; PRT; 515 AA.
AC Q8T108;
DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2002, sequence version 1.
```

```
DT 07-FEB-2006, entry version 10.
DE Heparanase-like protein.
GN Name=Bnhepa;
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
OC Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ps50; TISSUE=Posterior silkland;
RA Koike Y., Mita K., Suzuki M.G., Maeda S., Abe H., Osoegawa K.,
deJong P.J., Shimada T.;
RT "Genomic sequence of a 320-kb segment of the Z chromosome of Bombyx
mori containing a kettin ortholog.";
RL Mol. Genet. Genomics 269:137-149(2003).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB079860; BAB95191.1; -; Genomic DNA.
DR EMBL; AB090307; BAC10612.1; -; Genomic DNA.
DR InterPro; IPR005199; Glyco_hydro_79_N.
DR Pfam; PF03662; Glyco_hydro_79n; 1.
SQ SEQUENCE 515 AA; 59770 MW; FB8100ABE6EDDADB CRC64;

Query Match 46.2%; Score 48; DB 2; Length 515;
Best Local Similarity 62.5%; Pred. No. 38;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PAYLRFQGTGKTDLFIF 16
Db 84 PARLRLGGTMSERLIF 99

RESULT 34
Q47YDS_COLP3 PRELIMINARY; PRT; 638 AA.
AC Q47YDS;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2005, entry version 1.
DT 07-FEB-2006, entry version 3.
DE Prolyl oligopeptidase family protein (EC 3.4.21.26).
GN OrderedLocusNames=CPS_3511;
OS Colwellia psychrerythraea (strain 34H / ATCC BAA-681) (Vibrio
psychroerythrus).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Colwelliaceae; Colwellia.
OX NCBI_TaxID=167879;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=16043709; DOI=10.1073/pnas.0504766102;
RA Meche B.A., Nelson K.E., Deming J.W., Momen B., Melamud E., Zhang X.,
Moult J., Madupu R., Nelson W.C., Dodson R.J., Brinkac L.M.,
Daghighi S.C., Durkin A.S., DeBoy R.T., Kolonay J.F., Sullivan S.A.,
Zhou L., Daviden T.M., Wu M., Huston A.L., Lewis M., Weaver B.,
Weidman J.F., Khouri H., Uterback T.R., Feldblyum T.V., Fraser C.M.;
RT "The psychrophilic lifestyle as revealed by the genome sequence of
Colwellia psychrerythraea 34H through genomic and proteomic
analyses.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:10913-10918(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000083; AAZ24771.1; -; Genomic DNA.
DR TIGR; CPS 3511; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004287; F:prolyl oligopeptidase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis; IEA.
DR InterPro; IPR00379; Ser esters.
DR Pfam; PF00326; Peptidase_S9; 1.

KW Complete proteome; Hydrolase.
SQ SEQUENCE 638 AA; 71717 MW; 59DDA8B6C3BE4821 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 638;
Best Local Similarity 57.1%; Pred. No. 49;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 4 LREGGTTKTDLFIFD 17
Db 598 IRSGGTVDYLVFD 611

RESULT 35
Q8KZS8_ACEPA PRELIMINARY; PRT; 742 AA.
AC Q8KZS8;
DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2002, sequence version 1.
DT 07-FEB-2006, entry version 16.
DE Alcohol dehydrogenase.
OS Acetobacter pasteurianus (Acetobacter turbidans).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Acetobacteraceae; Acetobacter.
OX NCBI_TaxID=438;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCI 1193;
RA Takakuwa N., Yamane K., Oda Y., Fukaya M., Tsukamoto Y., Ohnishi M.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB086012; BAB97167.1; -; Genomic DNA.
DR HSSP; Q46444; 1KB0.
DR GO; GO:0030288; C:periplasmic space (sensu Gram-negative Bact. .; IEA.
DR GO; GO:0020037; F:heme binding; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR009056; Cyt c monohaem.
DR InterPro; IPR012282; Cytochrome_c_R.
DR InterPro; IPR001479; PQQ_bac.
DR InterPro; IPR002372; PQQ_repeat.
DR Pfam; PF01011; PQQ; 3.
DR SMART; SM00564; PQQ; 2.
DR PROSITE; PS00363; BACTERIAL_PQQ_1; 1.
DR PROSITE; PS51007; CYTC; 1.
SQ SEQUENCE 742 AA; 81802 MW; 42A7A50DCE2D981 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 742;
Best Local Similarity 66.7%; Pred. No. 58;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 7 GGTGKTDLFIFDP 18
Db 276 GGTVDLFVYDP 287

RESULT 36
Q2IRUO_RHOPA PRELIMINARY; PRT; 838 AA.
AC Q2IRUO;
DT 07-MAR-2006, integrated into UniProtKB/TrEMBL.
DT 07-MAR-2006, sequence version 1.
DT 07-MAR-2006, entry version 1.
DE Periplasmic sensor signal transduction histidine kinase precursor.
GN ORFNames=RPB_4383;
OS Rhodopseudomonas palustris Haa2.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Rhodopsedomonas.
OX NCBI_TaxID=316058;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Haa2;
```

```
RG US DOE Joint Genome Institute;
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Chain P., Malfatti S., Shin M.,
RA Vergez L., Schmutz J., Larimer F., Land M., Hauser L., Pelletier D.A.,
RA Kyrpides N., Anderson I., Oda Y., Harwood C.S., Richardson P.;
RT "Complete sequence of Rhodopseudomonas palustris Ha22.";
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; CP000250; ABD09070.1; --; Genomic_DNA.
KW Kinase; Signal.
FT SIGNAL 1 25 Potential.
SQ SEQUENCE 838 AA; 90636 MW; 0D237021D5ECB3C6 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 838;
Best Local Similarity 64.3%; Pred. No. 66;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 6 FGGTKTDFLIFDPK 19
DB 315 FGGTGRFTIFDPR 328
|||||
|||||

RESULT 37
Q3QRL3_9RHOB PRELIMINARY; PRT; 1895 AA.
ID Q3QRL3_9RHOB PRELIMINARY; PRT; 1895 AA.
AC Q3QRL3;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Outer membrane autotransporter barrel.
GN ORFNames=RoseDRAFT_1391;
OS Silicibacter sp. TW1040.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Silicibacter.
OX NCBI_TaxID=292414;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=TW1040;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Silicibacter sp.
TW1040.";
RT Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
RL [2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=TW1040;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Silicibacter sp. TW1040.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=TW1040;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AAFG0200006; EAN56260.1; --; Genomic_DNA.
DR InterPro; IPR006034; Asp/Glutamase.
DR InterPro; IPR005546; Auto transportbeta.
DR InterPro; IPR006315; Autotransporter.
DR InterPro; IPR001298; Filamin.
```

```
DR InterPro; IPR001969; Pept_Asp_AS.
DR Pfam; PF03797; Autotransporter; 1.
DR TIGFAMS; TIGR01414; autotrans_bar1; 1.
DR PROSITE; PS00144; ASN GLN ASE 1; UNKNOWN 3.
DR PROSITE; PS00141; ASP PROTEASE; UNKNOWN 1.
DR PROSITE; PS00194; FILAMIN REPEAT; 2.
SQ SEQUENCE 1895 AA; 188024 MW; 2A2C185A0FBD7866 CRC64;

Query Match 46.2%; Score 48; DB 2; Length 1895;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 7 GGTKTDFLIFDP 18
DB 101 GGSSTNFFIFDP 112
|||||
|||||

RESULT 38
Q2LZGO_DROPS PRELIMINARY; PRT; 2405 AA.
ID Q2LZGO_DROPS PRELIMINARY; PRT; 2405 AA.
AC Q2LZGO;
DT 21-FEB-2006, integrated into UniProtKB/TrEMBL.
DT 21-FEB-2006, sequence version 1.
DT 21-FEB-2006, entry version 1.
DE GA19954-PA (Fragment).
GN Name=Dpse\GA19954; ORFNames=Dpse GA19954;
OS Drosophila pseudoobscura (Fruit Fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7237;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MV2-25;
RX PubMed=1532085; DOI=10.1101/gr.3059305;
RA Richards S., Liu Y., Bettencourt B.R., Hradecky P., Letovsky S.,
RA Nielsen R., Thornton K., Hubisz M.J., Chen R., Meisel R.P.,
RA Couronne O., Hua S., Smith M.A., Zhang P., Liu J., Bussemaker H.J.,
RA van Batenburg M.F., Howells S.L., Scherer S.E., Sodergren E.,
RA Matthews B.B., Crosby M.A., Schroeder A.J., Ortiz-Barrientos D.,
RA Rives C.M., Metzker M.L., Muzny D.M., Scott G., Steffen D.,
RA Wheeler D.A., Worley K.C., Havlak P., Durbin K.J., Egan A., Gill R.,
RA Hume J., Morgan M.B., Miner G., Hamilton C., Huang Y., Waldron L.,
RA Verduzco D., Clerc-Blankenburg K.P., Dubchak I., Noor M.A.F.,
RA Anderson W., White K.P., Clark A.G., Schaeffer S.W., Gelbart W.,
RA Weinstock G.M., Gibbs R.A.;
RT "Comparative genome sequencing of Drosophila pseudoobscura:
chromosomal, gene, and cis-element evolution.";
RL Genome Res. 15:1-18(2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MV2-25;
RG FlyBase;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MV2-25;
RG Human Genome Sequencing Center;
RA Richards S., Liu Y., Bettencourt B.R., Hradecky P., Letovsky S.,
RA Chen R., Smith M.A., Howells S.L., Scherer S.E., Sodergren E.,
RA Rives C.M., Metzker M.L., Munz D.M., Wheeler D.A., Worley K.C.,
RA Havlak P., Durbin K.J., Egan A., Gill R., Hume J., Morgan M.B.,
RA Huang Y., Waldron L., Verduzco D., Blankenburg K.P., Adams C.,
RA Allen C., Allen H., Anyalebechi V., Asomugha C., Bellard T.,
RA Bhuchar V., Biswal K., Blair J., Blomstrom D., Burrell K.,
RA Calderon E., Cardenas V., Carter K., Cavazos I., Ceasar H., Chacko J.,
RA Chavez D., Chu J., Cockrell R., Cox C., Coyle M., Davila M., Davis C.,
RA Davy-Carroll L., De A., Delgado O., Denson S., Deramo C., Dinh H.,
RA Eaves K., Escotto M., Eugene C., Falls T., Fernandez S., Flagg N.,
RA Forbes L., Garner T., Garza M., Ghose S., Grady M., Hamilton C.,
RA Hernandez J., Hines S., Hogues M., Hollins B., Idlebird D., Iino K.,
RA Jimenez A., Johnson B., Jolivet A., Kelly S., King L., Kisano H.,
RA Kovar C., Lebow H., Lee K., LeGall F., Lewis L., Li Z., London P.,
```

```
RA Lopez J., Lozado R., Malloy K., Martinez E., Mercadao C., Miner G.,
RA Minja E., Moore S., Nanavati A., Ngo R., Nguyen N., Nwaokelenah O.,
RA Okwundu G., Parks K., Pasternak S., Patel B., Paul H., Payne C.,
RA Poindexter A., Primus E., Pu L.-L., Puazo M., Quiroz J., Rabata D.,
RA Reigh R., Ruiz S., Sanders W., Sison I., Sorelle R., Taylor C.,
RA Taylor T., Thomas N., Trejos Z., Usmani K., Vera V., Villanueva D.,
RA Wang S., Warren J., Warren R., White F., Wleczky R., Wright R.,
RA Noor M.A.F., Schaeffer S.W., Gelbart W., Weinstock G.M., Gibbs R.A.,
RA Weinstock G., Gibbs R.; to the EMBL/GenBank/DBJ databases.
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CH379069; EAL29548.1; -; Genomic_DNA.
FT NON TER 1
FT TER 2405 2405
SQ SEQUENCE 2405 AA; 256997 MW; 0456101FDB2A668B CRC64;

Query Match 46.2%; Score 48; DB 2; Length 2405;
Best Local Similarity 52.6%; Pred. No. 2.1e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 1 PAYLRFQGTGKTDFLIFDPK 19
Db 2138 PASLYFSQTDLLUSEPK 2156

RESULT 39
Q9WZB4.THEMA PRELIMINARY; PRT; 176 AA.
AC Q9WZB4;
DT 01-NOV-1999, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1999, sequence version 1.
DE Hypothetical protein.
GN OrderedLocNames=TM0646; ORFNames=TM_0646;
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
[1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=99287316; PubMed=10360571; DOI=10.1038/20601;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L.A., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.L.,
RA Heidelberg J.F., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB000512; AAD35730.1; -; Genomic_DNA.
DR PIR; E72351; E72351.
DR TIGR; TM0646; -.
DR BioCyc; TWAR2336:TM0646-MONOMER; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 176 AA; 2195 MW; D5768886C85A08FB CRC64;

Query Match 45.2%; Score 47; DB 2; Length 176;
Best Local Similarity 52.9%; Pred. No. 17;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 1 PAYLRFQGTGKTDFLIFD 17
Db 111 PKDARFICTPVDVVFVD 127
```

```
RESULT 40
Q5ZIQ7.CHICK PRELIMINARY; PRT; 1076 AA.
ID Q5ZIQ7.CHICK PRELIMINARY; PRT; 1076 AA.
AC Q5ZIQ7;
DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.
DT 23-NOV-2004, sequence version 1.
DE Hypothetical protein.
DE Hypothetical protein.
GN ORFNames=RCUMB04_24e3;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CB; TISSUE=Bursa;
RA Caldwell R.B., Kierzek A.M., Arakawa H., Bezzubov Y., Zaim J.,
RA Fiedler P., Kutter S., Blagodatki A., Kostovska D., Koter M.,
RA Plachy J., Carninci P., Hayashizaki Y., Buerstedde J.M.;
RT "Full-length cDNAs from chicken bursal lymphocytes to facilitate
RT gene function analysis.";
RL Genome Biol. 6:R6-R6(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AJ720727; CAG32386.1; -; mRNA.
DR GO; GO:0005488; F:binding; IEA.
DR InterPro; IPR011989; ARM-like.
DR InterPro; IPR005612; CBF.
DR Pfam; PF03914; CBF; 1.
DR Hypothetical protein.
KW SEQUENCE 1076 AA; 122590 MW; 683AE5974F95A164 CRC64;

Query Match 45.2%; Score 47; DB 2; Length 1076;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 LRFGGTKTDFLI 15
Db 45 LRLGGTKQDFLM 56

RESULT 41
Q5ZMV4.CHICK PRELIMINARY; PRT; 1076 AA.
ID Q5ZMV4.CHICK PRELIMINARY; PRT; 1076 AA.
AC Q5ZMV4;
DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.
DT 23-NOV-2004, sequence version 1.
DE Hypothetical protein.
DE Hypothetical protein.
GN ORFNames=RCUMB04_1b8;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CB; TISSUE=Bursa;
RA Caldwell R.B., Kierzek A.M., Arakawa H., Bezzubov Y., Zaim J.,
RA Fiedler P., Kutter S., Blagodatki A., Kostovska D., Koter M.,
RA Plachy J., Carninci P., Hayashizaki Y., Buerstedde J.M.;
RT "Full-length cDNAs from chicken bursal lymphocytes to facilitate
RT gene function analysis.";
RL Genome Biol. 6:R6-R6(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AJ719280; CAG30939.1; -; mRNA.
DR GO; GO:0005488; F:binding; IEA.
```

```
DR InterPro; IPR011989; ARM-like.
DR InterPro; IPR005612; CBF.
DR Pfam; PF03914; CBF; 1.
KW Hypothetical protein.
SQ SEQUENCE 1076 AA; 122549 MW; 7752F6970B8DD267 CRC64;

Query Match 45.2%; Score 47; DB 2; Length 1076;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 LREGGKTDFLI 15
Db 45 LRLGGTKQDFLM 56

RESULT 42
Q33FF2 METHU PRELIMINARY; PRT; 180 AA.
AC Q33FF2;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Conserved protein.
GN ORFNames=MhndRAFT_0483;
OS Methanospirillum hungatei JF-1.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanomicrobiales;
OC Methanospirillaceae; Methanospirillum.
OX NCBI_TaxID=323259;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=JF-1;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RA "Sequencing of the draft genome and assembly of Methanospirillum
RT hungatei JF-1.";
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=JF-1;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Larimer F., Land M.;
RA "Annotation of the draft genome assembly of Methanospirillum
RT JF-1.";
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AALU01000044; EAP14488.1; -; Genomic DNA.
DR EMBL; AALU01000044; EAP14488.1; -; Genomic DNA.
SQ SEQUENCE 180 AA; 21279 MW; 260C336E54315975 CRC64;

Query Match 44.2%; Score 46; DB 2; Length 180;
Best Local Similarity 47.1%; Pred. No. 26;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLI 17
Db 118 PSDARFLGSPVDIFVD 134

RESULT 43
Q4CAE2 CROWT PRELIMINARY; PRT; 219 AA.
ID Q4CAE2_CROWT
AC Q4CAE2;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2005, sequence version 1.
DT 21-FEB-2006, entry version 4.
DE Lipote-protein ligase B.
GN ORFNames=CwatDRAFT_6612;
```

```
OS Crocospaera watsonii.
OC Bacteria; Cyanobacteria; Chroococcales; Crocospaera.
OX NCBI_TaxID=165597;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=WH 8501;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RA "Sequencing of the draft genome and assembly of Crocospaera watsonii
RT WH 8501.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=WH 8501;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RA "Annotation of the draft genome assembly of Crocospaera watsonii WH
RT 8501.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=WH 8501;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RA Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AADV02000001; EAM52994.1; -; Genomic DNA.
DR GO; GO:0008445; F:acyltransferase activity; IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009107; P:lipoate biosynthesis; IEA.
DR GO; GO:0006464; P:protein modification; IEA.
DR InterPro; IPR000544; Lipoate_B.
DR Pfam; PF03099; BFL_LipoA_LipB; 1.
DR Prodom; PD006086; Lipoate_B; 1.
DR TIGRFAMs; TIGR00214; LipB; 1.
DR PROSITE; PS01313; LipB; 1.
KW Ligase.
SQ SEQUENCE 219 AA; 25080 MW; F44134A46CF65B9D CRC64;

Query Match 44.2%; Score 46; DB 2; Length 219;
Best Local Similarity 47.4%; Pred. No. 33;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 PAYLRFGGTKTDFLI 19
Db 53 PVTYLTGSGSTKFKIFDPK 71

RESULT 44
Q5TMQ4 ANOGA PRELIMINARY; PRT; 308 AA.
ID Q5TMQ4_ANOGA
AC Q5TMQ4;
DT 07-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 07-DEC-2004, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE ENSANGP00000026025.
GN ORFNames=ENSANGG00000022402;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP NUCLEOTIDE SEQUENCE.
```

```
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAAB01008986; EAL38841.1; -; Genomic DNA.
DR GO; GO:0001175; F:3'-5'-exoribonuclease activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0006336; F:RNA processing; IEA.
DR InterPro; IPR001247; ExorNase.
DR Pfam; PF01138; RNase PH; 1.
SQ SEQUENCE 308 AA; 33784 MW; 3A2F1F2B872F8FA7 CRC64;

Query Match 44.2%; Score 46; DB 2; Length 308;
Best Local Similarity 44.4%; Pred. No. 48;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 2 AYLRFGGKTDFLIIFDPK 19
Db 69 AYVGLGNTKIVSVFDPK 86

RESULT 45
Q2W843 MAGSA
ID Q2W843 MAGSA PRELIMINARY; PRT; 319 AA.
AC Q2W843;
DT 10-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 10-JAN-2006, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Hypothetical protein.
GN ORFNames=amb1178;
OS Magnetospirillum magneticum AMB-1.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Rhodospirillaceae; Magnetospirillum.
OX NCBI_TaxID=342108;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AMB-1;
RX PubMed=16303747;
RA Takeyama H.;
RT "Complete Genome Sequence of the Facultative Anaerobic Magnetotactic
Bacterium Magnetospirillum sp. strain AMB-1.";
RL DNA Res. 12:157-166(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AP007255; BAE49982.1; -; Genomic DNA.
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 319 AA; 33449 MW; FDA8F6475858BE5D CRC64;

Query Match 44.2%; Score 46; DB 2; Length 319;
Best Local Similarity 52.9%; Pred. No. 50;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 PAYLRFGGKTDFLIIFD 17
Db 126 PKFFRHGGAKLDKLSFD 142

RESULT 46
Q2W843 MAGSA
ID Q2W843 MAGSA PRELIMINARY; PRT; 319 AA.
AC Q2W843;
DT 10-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 10-JAN-2006, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Hypothetical protein.
GN ORFNames=amb1178;
OS Magnetospirillum magneticum AMB-1.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Rhodospirillaceae; Magnetospirillum.
OX NCBI_TaxID=342108;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AMB-1;
RX PubMed=16303747;
RA Takeyama H.;
RT "Complete Genome Sequence of the Facultative Anaerobic Magnetotactic
Bacterium Magnetospirillum sp. strain AMB-1.";
RL DNA Res. 12:157-166(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AP007255; BAE49982.1; -; Genomic DNA.
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 319 AA; 33449 MW; FDA8F6475858BE5D CRC64;

Query Match 44.2%; Score 46; DB 2; Length 319;
Best Local Similarity 52.9%; Pred. No. 50;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 PAYLRFGGKTDFLIIFD 17
Db 126 PKFFRHGGAKLDKLSFD 142

RESULT 47
Q57MP2 SALCH
ID Q57MP2 SALCH PRELIMINARY; PRT; 404 AA.
AC Q57MP2;
DT 10-MAY-2005, integrated into UniProtKB/TrEMBL.
DT 10-MAY-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Putative colanic acid polymerase.
GN Name=wcaD; OrderedLocustNames=SC2113; ORFNames=SCH_2113;
OS Salmonella choleraesuis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=591;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=SC-B67;
RX PubMed=15781495; DOI=10.1093/nar/gki297;
RA Chiu C.-H., Tang P., Chu C., Hu S., Bao Q., Yu J., Chou Y.-Y.,
Wang H.-S., Lee Y.-S.;
RT "The genome sequence of Salmonella enterica serovar Choleraesuis, a
highly invasive and resistant zoonotic pathogen.";
RL Nucleic Acids Res. 33:1690-1698(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB017220; AAX66019.1; -; Genomic DNA.
KW Complete proteome.
SQ SEQUENCE 404 AA; 45129 MW; AB77A8DCE1812351 CRC64;

Query Match 44.2%; Score 46; DB 2; Length 404;
Best Local Similarity 43.8%; Pred. No. 65;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 3 YLRFGGKTDFLIIFD 18
```

```
Q2WAD9 MAGSA
ID Q2WAD9 MAGSA PRELIMINARY; PRT; 319 AA.
AC Q2WAD9;
DT 10-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 10-JAN-2006, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Hypothetical protein.
GN ORFNames=amb0382;
OS Magnetospirillum magneticum AMB-1.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Rhodospirillaceae; Magnetospirillum.
OX NCBI_TaxID=342108;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AMB-1;
RX PubMed=16303747;
RA Matsunaga T., Okamura Y., Fukuda Y., Wahyudi A.T., Murase Y.,
Takeyama H.;
RT "Complete Genome Sequence of the Facultative Anaerobic Magnetotactic
Bacterium Magnetospirillum sp. strain AMB-1.";
RL DNA Res. 12:157-166(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AP007255; BAE49186.1; -; Genomic DNA.
KW Hypothetical protein.
SQ SEQUENCE 319 AA; 33463 MW; FCD091595A652FC3 CRC64;

Query Match 44.2%; Score 46; DB 2; Length 319;
Best Local Similarity 52.9%; Pred. No. 50;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 PAYLRFGGKTDFLIIFD 17
Db 126 PKFFRHGGAKLDKLSFD 142

RESULT 47
Q57MP2 SALCH
ID Q57MP2 SALCH PRELIMINARY; PRT; 404 AA.
AC Q57MP2;
DT 10-MAY-2005, integrated into UniProtKB/TrEMBL.
DT 10-MAY-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Putative colanic acid polymerase.
GN Name=wcaD; OrderedLocustNames=SC2113; ORFNames=SCH_2113;
OS Salmonella choleraesuis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=591;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=SC-B67;
RX PubMed=15781495; DOI=10.1093/nar/gki297;
RA Chiu C.-H., Tang P., Chu C., Hu S., Bao Q., Yu J., Chou Y.-Y.,
Wang H.-S., Lee Y.-S.;
RT "The genome sequence of Salmonella enterica serovar Choleraesuis, a
highly invasive and resistant zoonotic pathogen.";
RL Nucleic Acids Res. 33:1690-1698(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB017220; AAX66019.1; -; Genomic DNA.
KW Complete proteome.
SQ SEQUENCE 404 AA; 45129 MW; AB77A8DCE1812351 CRC64;

Query Match 44.2%; Score 46; DB 2; Length 404;
Best Local Similarity 43.8%; Pred. No. 65;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 3 YLRFGGKTDFLIIFD 18
```



```
Db 162 FIRPGGKRTTALYFEP 177
:::||||:| | | | |
RESULT 48
QSPDY6_SALPA PRELIMINARY; PRT; 404 AA.
AC QSPDY6;
DT 04-JAN-2005, integrated into UniProtKB/TrEMBL.
DT 04-JAN-2005, sequence version 1.
DT 07-FEB-2006, entry version 7.
DE Putative colanic acid polymerase.
GN Name=wcaD; OrderedLocusNames=SPA0754;
OS Salmonella paratyphi-a.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=54388;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 9150 / SARB42;
RX PubMed=15311882; DOI=10.1038/ng1470;
RA McClelland M., Sanderson K.E., Clifton S.W., Latreille P.,
RA Porwollik S., Sabo A., Meyer R., Bieri T., Ozersky P., McEllan M.,
RA Harkins C.R., Wang C., Nguyen C., Berghoff A., Elliott G.,
RA Kohlberg S., Strong C., Du F., Carter J., Kremizki C., Layman D.,
RA Leonard S., Sun H., Fulton L., Nash W., Miner T., Minx P.,
RA Dalehaunt K., Fronick C., Magrini V., Nhan M., Warren W., Florea L.,
RA Spieth J., Wilson R.K.;
RT "Comparison of genome degradation in Paratyphi A and Typhi, human-
RT restricted serovars of Salmonella enterica that cause typhoid.";
RL Nat. Genet. 36:1268-1274(2004).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DE EMBL; CP000026; AAU76750.1; -; Genomic_DNA.
KW Complete proteome.
SQ SEQUENCE 404 AA; 45129 MW; AB77A8DC1812351 CRC64;

Query Match 44.2%; Score 46; DB 2; Length 404;
Best Local Similarity 43.8%; Pred. No. 65;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 3 YLRFGGTTKDFLI FDP 18
:::||||:| | | | |
Db 162 FIRPGGKRTTALYFEP 177

RESULT 49
Q9F7A7_SALTY PRELIMINARY; PRT; 404 AA.
AC Q9F7A7; Q7CQA2;
DT 01-MAR-2001, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2001, sequence version 1.
DT 07-FEB-2006, entry version 14.
DE Hypothetical protein wcaD (Putative colanic acid polymerase).
GN Name=wcaD; OrderedLocusNames=STM2112;
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=LT2;
RX PubMed=20461159; PubMed=11004393;
RA Stevenson G., Lan R., Reeves P.R.;
RT "The colanic acid gene cluster of Salmonella enterica has a complex
RT history.";
RL FEMS Microbiol. Lett. 191:11-16(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;
```

```
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DE EMBL; AF285084; AAG24810.1; -; Genomic_DNA.
DR EMBL; AE008793; AAL21016.1; -; Genomic_DNA.
DR BioCyc; STYV99287:STM2112-MONOMER; -.
DR LinkHub; Q9F7A7; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 404 AA; 45147 MW; AA72F888E36D0D91 CRC64;

Query Match 44.2%; Score 46; DB 2; Length 404;
Best Local Similarity 43.8%; Pred. No. 65;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 3 YLRFGGTTKDFLI FDP 18
:::||||:| | | | |
Db 162 FIRPGGKRTTALYFEP 177

RESULT 50
WCAD_ECOLI STANDARD; PRT; 405 AA.
ID WCAD_ECOLI
AC P71238; P76385;
DT 01-NOV-1997, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1997, sequence version 2.
DT 07-NOV-2006, entry version 34.
DE Putative colanic acid polymerase.
GN Name=wcaD; OrderedLocusNames=b2056;
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=K12;
RX MEDLINE=96326333; PubMed=8759852;
RA Stevenson G., Adrianopoulos K., Hobbs M., Reeves P.R.;
RT "Organization of the Escherichia coli K-12 gene cluster responsible
RT for production of the extracellular polysaccharide colanic acid.";
RL J. Bacteriol. 178:4885-4893(1996).
RN [2]
RP SEQUENCE REVISION.
RC STRAIN=K12;
RA Reeves P.R.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SUBCELLULAR LOCATION.
RC STRAIN=K12 / MG1655;
RX PubMed=15919996; DOI=10.1126/science.1109730;
RA Daley D.O., Rapp M., Granseth E., Melen K., Drew D., von Heijne G.;
RT "Global topology analysis of the Escherichia coli inner membrane
RT proteome.";
RL Science 308:1321-1323(2005).
CC -!- PATHWAY: Slime polysaccharide colanic acid biosynthesis.
```



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:21 ; Search time 129.765 Seconds  
(without alignments)  
3728.138 Million cell updates/sec

Title: US-10-645-659A-5  
Perfect score: 2728  
Sequence: 1 MVLVLLVLLVAVPPRTAE.....LPAFSYGFYVIRNAKAIACI 523

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 7.2.\*  
1: uniprot\_prot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2728	100.0	523	1 HPSE_CHICK	Q90YK5 gallus gall
2	1653	60.6	574	2 Q33X66_SPAJJD	Q33X66 spalax juda
3	1649	60.4	574	2 Q33X77_9RODE	Q33X77 spalax carm
4	1648	60.4	543	1 HPSE_HUMAN	Q9Y251 homo sapien
5	1641	60.2	574	2 Q33X79_9RODE	Q33X79 spalax gali
6	1640	60.1	574	2 Q33X88_9RODE	Q33X88 spalax gola
7	1624	59.5	545	1 HPSE_BOVIN	Q9MY90 bos taurus
8	1602	58.7	535	1 HPSE_MOUSE	Q6Y9Z1 mus musculus
9	1572	57.6	536	1 HPSE_RAT	Q71P11 rattus norv
10	1558	57.1	558	2 Q33X55_SPAJJD	Q33X55 spalax juda
11	1232	45.2	533	2 Q4SYF6_TETNG	Q4SYF6 tetraodon n
12	1028.5	37.7	592	1 HPSE2_HUMAN	Q8WQZ2 homo sapien
13	1028.5	37.7	592	2 Q2MLH9_HUMAN	Q2MLH9 homo sapien
14	936.5	34.3	597	2 Q4TB80_TETNG	Q4TB80 tetraodon n
15	679	24.9	255	2 Q4T8C8_TETNG	Q4T8C8 tetraodon n
16	640	23.5	515	2 Q8T108_BOMMO	Q8T108 bombyx mori
17	395.5	14.5	539	2 Q2QN56_ORYSA	Q2QN56 oryza sativ
18	395	14.5	543	1 HPSE1_ARATH	Q9FF10 arabidopsis
19	387.5	14.2	529	2 Q6ZJE2_ORYSA	Q6ZJE2 oryza sativ
20	382	14.0	526	2 Q5SNA6_ORYSA	Q5SNA6 oryza sativ
21	375.5	13.8	537	2 Q70YJ3_HORVU	Q70YJ3 hordeum vul
22	351.5	12.9	541	2 Q69116_ORYSA	Q69116 oryza sativ
23	350.5	12.8	536	1 HPSE3_ARATH	Q9FZP1 arabidopsis
24	346	12.7	527	2 Q9LRC8_SCUBA	Q9LRC8 scutellaria
25	339	12.4	539	1 HPSE2_ARATH	Q81608 arabidopsis
26	336.5	12.3	401	2 Q30324_ARATH	Q30324 arabidopsis
27	336.5	12.3	559	2 Q89F99_BRAJA	Q89F99 bradyrhizob
28	317.5	11.6	516	2 Q447R5_SOLUS	Q447R5 solibacter
29	289	10.6	506	2 Q37Q70_SPHAR	Q37Q70 novosphingo
30	257	9.4	382	2 Q3EBP7_ARATH	Q3EBP7 arabidopsis
31	249.5	9.1	537	2 Q43803_SOLUS	Q43803 solibacter

32	186.5	6.8	463	2	Q63T97_BURPS	Q63T97 burkholderi
33	185.5	6.8	670	2	Q3JTG0_BURP1	Q3JTG0 burkholderi
34	139.5	5.1	1128	2	Q5TT65_ANOGA	Q5TT65 anopheles g
35	134.5	4.9	510	2	Q2U0T3_ASPOR	Q2U0T3 aspergillus
36	129	4.7	935	2	Q9VE79_DROME	Q9VE79 drosophila
37	127	4.7	484	2	Q5RFE6_PONPY	Q5RFE6 pongo pygma
38	122.5	4.5	533	2	Q7N6A1_PHOLL	Q7N6A1 photorhabdu
39	120	4.4	559	2	Q7SFB0_NEUCR	Q7SFB0 neuropeptora
40	120	4.4	597	2	Q62EB8_BURMA	Q62EB8 burkholderi
41	120	4.4	597	2	Q631W2_BURPS	Q631W2 burkholderi
42	120	4.4	642	2	Q3JJN8_BURP1	Q3JJN8 burkholderi
43	119.5	4.4	533	2	Q7N666_PHOLL	Q7N666 photorhabdu
44	119	4.4	1630	2	Q5B9V9_EMENI	Q5B9V9 aspergillus
45	117.5	4.3	373	2	Q39AY5_BURS3	Q39AY5 burkholderi

ALIGNMENTS

RESULT 1  
HPSE\_CHICK ID HPSE\_CHICK STANDARD; PRT; 523 AA.  
AC Q90YK5;  
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
DT 01-DEC-2001, sequence version 1.  
DT 07-FEB-2006, entry version 12.  
DE Heparanase precursor (EC 3.2.-.-).  
GN Name=HPSE; Synonyms=HPA;  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP MEDLINE=2136959; PubMed=11387326; DOI=10.1074/jbc.M102462200;  
RA Goldshmidt O., Zoharia E., Aingorn H., Guatta-Rangini Z., Atmon R.,  
RA Michal I., Pecker I., Mitrani E., Vlodavsky I.;  
RT "Expression pattern and secretion of human and chicken heparanase are  
RT determined by their signal peptide sequence.";  
J. Biol. Chem. 276:29178-29187(2001).  
RL  
CC -!- FUNCTION: Endoglycosidase which is a cell surface and  
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
CC proteoglycans (By similarity).  
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted  
CC (By similarity).  
CC -!- PTM: N-glycosylated (By similarity).  
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonCommercial License  
CC  
CC EMBL; AY037007; AA82648.1; -; mRNA.  
CC Ensembl; ENSGALG00000011203; Gallus gallus.  
CC InterPro; IPR005199; Glyco\_hydro\_79\_N.  
CC Pfam; PF03662; Glyco\_hydro\_79n; 1.  
CC KMW Glycoprotein; Hydrolase; Lysosome; Membrane; Signal.  
FT SIGNAL 1 18  
FT CHAIN 19 523 Heparanase.  
FT /FTid=PRO\_0000042259.  
FT REGION 137 141 Heparin/HS-binding (By similarity).  
FT REGION 250 260 Heparin/HS-binding (By similarity).  
FT ACT\_SITE 204 204 Proton donor (Potential).  
FT ACT\_SITE 323 323 Nucleophile (Potential).  
FT CARBOHYD 141 141 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 196 196 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 436 436 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 439 439 N-linked (GlcNAc...) (Potential).  
SQ SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF881 CRC64;

Query Match 100.0%; Score 2728; DB 1; Length 523;  
Best Local Similarity 100.0%; Pred. No. 5.6e-207;

```
Matches 523; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MLVLLVLLVLLVPPRTAELQGLREPICAVSPAFSLTLDASLARDPRFVALLRHPKL 60
Db 1 MLVLLVLLVLLVPPRTAELQGLREPICAVSPAFSLTLDASLARDPRFVALLRHPKL 60
Qy 61 HTLASGLSPGLRFGCTSTDFLIFNPKNKDSWEEKVLSFQAKDVCEAWPSPFAVVPKLL 120
Db 61 HTLASGLSPGLRFGCTSTDFLIFNPKNKDSWEEKVLSFQAKDVCEAWPSPFAVVPKLL 120
Qy 121 TOWPLOEKLLLAHSHKWKHKNTTITRSTLDILHTFASSSGFRLVFGLNALLRRAGLOWDS 180
Db 121 TOWPLOEKLLLAHSHKWKHKNTTITRSTLDILHTFASSSGFRLVFGLNALLRRAGLOWDS 180
Qy 181 SNAKQLLGCAQRSYNI SWELGNPNRSFRKKSIGICIDGFGOLGRDFVHLRQLLSQHPLRYH 240
Db 181 SNAKQLLGCAQRSYNI SWELGNPNRSFRKKSIGICIDGFGOLGRDFVHLRQLLSQHPLRYH 240
Qy 241 AELYLGLDVGPQRKHTQHLRLSRFMSKGGKAIDSVTWHYYVNGRSATREDFLSPVLDSPA 300
Db 241 AELYLGLDVGPQRKHTQHLRLSRFMSKGGKAIDSVTWHYYVNGRSATREDFLSPVLDSPA 300
Qy 301 TAIHDLVGLIVEATVPCKKWLGETGSAYGGAPQLSNTYVAGFMWLDKGLAARRGIDVV 360
Db 301 TAIHDLVGLIVEATVPCKKWLGETGSAYGGAPQLSNTYVAGFMWLDKGLAARRGIDVV 360
Qy 361 MRQVSEFAGSYHLVDAGFKPLPDYWLISLYKRLVGTQVLSQASVEQADARRPRVYLHCTNP 420
Db 361 MRQVSEFAGSYHLVDAGFKPLPDYWLISLYKRLVGTQVLSQASVEQADARRPRVYLHCTNP 420
Qy 421 RHPKYREGDVTLPALNLSNVTQSLQPKQWLSKSDVQYLLPHGKDSILSREVQLNGRL 480
Db 421 RHPKYREGDVTLPALNLSNVTQSLQPKQWLSKSDVQYLLPHGKDSILSREVQLNGRL 480
Qy 481 QMVDDETLPALHEMALAPGSTGLPFAFSYGFYVIRNAKAIACI 523
Db 481 QMVDDETLPALHEMALAPGSTGLPFAFSYGFYVIRNAKAIACI 523

RESULT 2 SPAJD
Q333X6 SPAJD PRELIMINARY; PRT; 574 AA.
AC Q333X6;
DT 06-DEC-2005, integrated into UniProtKB/TREMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax judaei (Blind subterranean mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=134510;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166 (2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0 (0).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; AM085493; CAJ30020.1; -; mRNA.
SQ SEQUENCE 574 AA; 64515 MW; 3AE8B13F07451684 CRC64;
```

```
Query Match 60.4%; Score 1649; DB 2; Length 574;
Best Local Similarity 60.7%; Pred. No. 1.7e-121;
Matches 315; Conservative 85; Mismatches 117; Indels 2; Gaps 2;
```

```
Query Match 60.6%; Score 1653; DB 2; Length 574;
Best Local Similarity 60.7%; Pred. No. 8.4e-122;
Matches 315; Conservative 86; Mismatches 116; Indels 2; Gaps 2;
Qy 6 LLVLLVPPRTAELQGLREPICAVSPAFSLTLDASLARDPRFVALLRHPKLHTLAS 65
Db 57 LVQCILAAQADVVELEFSTQRPDLHLVSPSFLSITIDANLATDPRFLTFLSPKDALAR 116
Qy 66 GLSPGFLRFGCTSTDFLIFNPKNKDSWEEKVLSFQAKDVCEAWPSPFAVVPKLLTOWP 124
Db 117 GLSPAYLRFGGTKTDFLIFDPKKEPSHEERSYWKSVQVNHDCRSAGIAPVVVRRLQVEMP 176
Qy 125 LOEKLLLAHSHKWKHKNTTITRSTLDILHTFASSSGFRLVFGLNALLRRAGLOWDSNAK 184
Db 177 FQEQLLIREQYQKEFKNSYSSRSDMLYTFARCSGLDLIFGLNALLRTADFRWNSSNAQ 236
Qy 185 QLLOYCAQRSYNI SWELGNPNRSFRKKSIGICIDGFGOLGRDFVHLRQLLSQHPLRYHAEL 244
Db 237 LLLNYCSSKNYDI SWELGNPNRSFRKKAHISIDGLQLGEDYIELRKLKSTL-KNVKLY 295
Qy 245 GLDVGPQRKHTQHLRLSRFMSKGGKAIDSVTWHYYVNGRSATREDFLSPVLDSPATAIH 304
Db 296 GPDVGQPRGKTVKLLRSFLKAGEVIDSVTWHYYVNGRIATKEDFLSPDVLDTFILSVQ 355
Qy 305 DVLGIVEATVPCKKWLGETGSAYGGAPQLSNTYVAGFMWLDKGLAARRGIDVVMRQV 364
Db 356 KILQVVEETRPCKKWLGETSSAYGGAPLLSNTFAAGFMWLDKGLSAQMGIEVVMRQV 415
Qy 365 SFAGSYHLVDAGFKPLPDYWLISLYKRLVGTQVLSQASVEQADARRPRVYLHCTNPRPK 424
Db 416 FFGAGNYHLVDKNEPEPLPDYWLISLLFKLVGSKVLMARVGPDRSKLRVYLHCTNINHR 475
Qy 425 YREGDVTLPALNLSNVTQSLQPKQWLSKSDVQYLLPHGKDSILSREVQLNGRLQWMD 484
Db 476 YQEGDLTYALNLYNVTKHKLQVLFKNKPVDKLVLPGLPGGLSKSVQLNGQALQWMD 535
Qy 485 DETLPALHEMALAPGSTGLPFAFSYGFYVIRNAKAIACI 523
Db 536 DQTLPALTEKPLRPGSSGLPFAFSYGFYVIRNAKVAACL 574

RESULT 3
Q333X7 9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X7;
DT 06-DEC-2005, integrated into UniProtKB/TREMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax carmeli.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164324;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166 (2005).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL; AM085492; CAJ30019.1; -; mRNA.
SQ SEQUENCE 574 AA; 64459 MW; 9F1D19DCB8BD99DE CRC64;
```

```
Query Match 60.4%; Score 1649; DB 2; Length 574;
Best Local Similarity 60.7%; Pred. No. 1.7e-121;
Matches 315; Conservative 85; Mismatches 117; Indels 2; Gaps 2;
```





OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muroidae; Spalacidae; Spalacinae; Spalax.  
 OX NCBI\_TaxID=164323;  
 RN NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Kidney;  
 RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
 RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
 cloning and identification of a novel splice variant.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).  
 RN NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Kidney;  
 RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
 RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
 cloning and identification of a novel splice variant.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).  
 CC -----  
 CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; AM085490; CAJ30017.1; -; mRNA.  
 SQ SEQUENCE 574 AA; 64525 MW; 1635865051B380D0 CRC64;  
 -----  
 Query Match 60.2%; Score 1641; DB 2; Length 574;  
 Best Local Similarity 60.5%; Pred. No. 7.5e-121;  
 Matches 314; Conservative 84; Mismatches 119; Indels 2; Gaps 2;  
 QY 6 LVLVLLAVPPRTAEQLGLREPIGAVSPAFSLTLDASLARDPRFVALLRHPKLUHTLAS 65  
 DB 57 LVQCILAAQAEVDVELEFSTQRLHLVSPSFLSIDANLATDPRFLTGLSPKLRALAR 116  
 QY 66 GLSPGFLRFQGTSTDFLIFPNKDKSTWEEKVSEFOAK-DVCEAWPSFAVVPKLLLTQWP 124  
 DB 117 GLSPAYLRFQGTSTDFLIFPKPEPSHEERSYWKSVQVNHIDCRSGAIPAVVVRRLQVWP 176  
 QY 125 LQEKLLAAHSWKHKNTTITRSTLDILHTFASSSGFRLVGLNALLRRAGLQWDSNAK 184  
 DB 177 FQEQLLREQYQKDFKNSTYSRSSVDMLYTFARCSGLDLIFGLNALLRTADFRWNSSNAQ 236  
 QY 185 QLLGYCAQRSYNTSWELGNPNFRKSGICIDGFLQGRDFVHLRQLLSQHPLRYHAELY 244  
 DB 237 LLLNYCCKNYDISWELGNPNFRKSGAHSIDGLQGEDYIELHKLKSTL-KNVKLY 295  
 QY 245 GLDVGPRKHTQHLRSFMKSGKAIDSVTHHYVNGRSATREDPLSPVLDSEFATAIH 304  
 DB 296 GPDVGQPRGKTVKLLRSFLKAGGEVDSVTWHYVNLNGRIATKEDFLSPDVLDTFLSVQ 355  
 QY 305 DVLGIVEATVPKKVWLGETSGAYGGAPQLSNTYVAGFMWLDKGLAARRGIDVVMRQV 364  
 DB 356 KILQVVEETRPKKVWLGETSSAYGGAPLLSNTFAAGFMWLDKGLSAGWGLEVMRQV 415  
 QY 365 SFGAGSYHLVDAGFKPLPDYWLSSLYKRLVGTTRVLQASVEQADARRPRVYLHCTNPRHPK 424  
 DB 416 FFGAGNYHLVDKNFEPPLDYWLSLLPKLVGSKVMARVGPDRSKLRYLHCTNINHRP 475  
 QY 425 YREGDVTFLALNLSNTQSLQPKLWSKVDQYLLPHGKDSITLSREVQLNGELQWVD 484  
 DB 476 YQEGDLTLALNLYNVTKHLKLPYQLFNKPDVLYVPLPGPGLLSKSVQLNGQALQWVD 535  
 QY 485 DETLPALHEMALAPGSTGLPAPFSYGFYVIRNAKAIACI 523  
 DB 536 DQTLPALTEKPLRPGSSGLPAPFSYGFYVIRNAKVAACL 574  
 -----  
 RESULT 6  
 Q333X8\_9RODE PRELIMINARY; PRT; 574 AA.  
 AC Q333X8;  
 DT 06-DEC-2005, integrated into UniprotKB/TrEMBL.  
 DT 06-DEC-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 3.

DE Heparanase.  
 GN Name=hpa.  
 OS Spalax golani.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muroidae; Spalacidae; Spalacinae; Spalax.  
 OX NCBI\_TaxID=191382;  
 RN NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Kidney;  
 RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
 RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
 cloning and identification of a novel splice variant.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).  
 RN NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Kidney;  
 RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;  
 RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
 cloning and identification of a novel splice variant.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).  
 CC -----  
 CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; AM085491; CAJ30018.1; -; mRNA.  
 SQ SEQUENCE 574 AA; 64555 MW; 48BEFECE7D0BCB34 CRC64;  
 -----  
 Query Match 60.1%; Score 1640; DB 2; Length 574;  
 Best Local Similarity 60.3%; Pred. No. 8.9e-121;  
 Matches 313; Conservative 86; Mismatches 118; Indels 2; Gaps 2;  
 QY 6 LVLVLLAVPPRTAEQLGLREPIGAVSPAFSLTLDASLARDPRFVALLRHPKLUHTLAS 65  
 DB 57 LVQCILAAQAEVDVELEFSTQRLHLVSPSFLSIDANLATDPRFLTGLSPKLRALAR 116  
 QY 66 GLSPGFLRFQGTSTDFLIFPNKDKSTWEEKVSEFOAK-DVCEAWPSFAVVPKLLLTQWP 124  
 DB 117 GLSPAYLRFQGTSTDFLIFPKPEPSHEERSYWKSVQVNHIDCRSGAIPAVVVRRLQVWP 176  
 QY 125 LQEKLLAAHSWKHKNTTITRSTLDILHTFASSSGFRLVGLNALLRRAGLQWDSNAK 184  
 DB 177 FQEQLLREQYQKDFKNSTYSRSSVDMLYTFARCSGLDLIFGLNALLRTADFRWNSSNAQ 236  
 QY 185 QLLGYCAQRSYNTSWELGNPNFRKSGICIDGFLQGRDFVHLRQLLSQHPLRYHAELY 244  
 DB 237 LLLNYCCKNYDISWELGNPNFRKSGAHSIDGLQGEDYIELHKLKSTL-KNVKLY 295  
 QY 245 GLDVGPRKHTQHLRSFMKSGKAIDSVTHHYVNGRSATREDPLSPVLDSEFATAIH 304  
 DB 296 GPDVGQPRGKTVKLLRSFLKAGGEVDSVTWHYVNLNGRIATKEDFLSPDVLDTFLSVQ 355  
 QY 305 DVLGIVEATVPKKVWLGETSGAYGGAPQLSNTYVAGFMWLDKGLAARRGIDVVMRQV 364  
 DB 356 KILQVVEETRPKKVWLGETSSAYGGAPLLSNTFAAGFMWLDKGLSAGWGLEVMRQV 415  
 QY 365 SFGAGSYHLVDAGFKPLPDYWLSSLYKRLVGTTRVLQASVEQADARRPRVYLHCTNPRHPK 424  
 DB 416 FFGAGNYHLVDKNFEPPLDYWLSLLPKLVGSKVMARVGPDRSKLRYLHCTNINHRP 475  
 QY 425 YREGDVTFLALNLSNTQSLQPKLWSKVDQYLLPHGKDSITLSREVQLNGELQWVD 484  
 DB 476 YQEGDLTLALNLYNVTKHLKLPYQLFNKPDVLYVPLPGPGLLSKSVQLNGQALQWVD 535  
 QY 485 DETLPALHEMALAPGSTGLPAPFSYGFYVIRNAKAIACI 523  
 DB 536 DQTLPALTEKPLRPGSSGLPAPFSYGFYVIRNAKVAACL 574  
 -----  
 RESULT 7  
 HPSE\_BOVIN  
 ID HPSE\_BOVIN  
 AC Q9MY0;  
 STANDARD; PRT; 545 AA.



DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
DT 01-JUN-2001, sequence version 2.  
DT 07-MAR-2006, entry version 13.  
DE Heparanase precursor (EC 3.2.-.-) [Contains: Heparanase 8 kDa subunit;  
DE Heparanase 50 kDa subunit].  
GN Name=HPSE;  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
OC Pecora; Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.  
RC TISSUE=Placenta;  
RX MEDLINE=2117669; PubMed=11277877;  
RA Kizaki K., Nakano H., Nakano H., Takahashi T., Imai K., Hashizume K.;  
RT "Expression of heparanase mRNA in bovine placenta during gestation.";  
RL Reproduction 121:573-580(2001).  
CC -!- FUNCTION: Endoglycosidase which is a cell surface and  
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
CC proteoglycans. Also implicated in the extravasation of leukocytes  
CC and tumor cell lines. Contributes to metastasis and angiogenesis  
CC (By similarity).  
CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower  
CC extent, by heparin, sulfamin and EDTA. Activated by calcium and  
CC magnesium (By similarity).  
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
CC subunits, the proteolytic products (By similarity).  
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.  
CC Secreted, internalised and transferred to late endosomes/lysosomes  
CC as a proheparanase. In lysosomes, it is processed into the active  
CC form, the heparanase. The uptake or internalisation of  
CC proheparanase is mediated by HSPGs. Heparin appears to be a  
CC competitor and retain proheparanase in the extracellular medium  
CC (By similarity).  
CC -!- TISSUE SPECIFICITY: Highly expressed in placenta and weakly in the  
CC kidney, lung, spleen and uterus.  
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form  
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa  
CC product. The active form, the 8/50 kDa heterodimer, is resistant  
CC to degradation. Complete removal of the linker peptide appears to  
CC be a prerequisite to the complete activation of the enzyme (By  
CC similarity).  
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears  
CC to be essential for its solubility (By similarity).  
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL: AF281160; AAF87301.2; -; mRNA.  
DR InterPro: IPR005199; Glyco\_hydro\_79\_N.  
DR Pfam: PF03662; Glyco\_Hydro\_79n; 1.  
KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;  
KW Signal.  
FT SIGNAL 1 37 By similarity.  
FT CHAIN 38 111 Heparanase 8 kDa subunit (By similarity).  
FT FT /FTid=PRO\_0000042256.  
FT PROPEP 112 159 Linker peptide.  
FT FT /FTid=PRO\_0000042257.  
FT CHAIN 160 545 Heparanase 50 kDa subunit (By  
FT similarity).  
FT FT /FTid=PRO\_0000042258.  
FT REGION 160 164 Heparin/HS-binding (Potential).  
FT REGION 272 282 Heparin/HS-binding (Potential).  
FT ACT\_SITE 227 227 Proton donor (Potential).  
FT ACT\_SITE 345 345 Nucleophile (Potential).  
FT CARBOHYD 164 164 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 219 219 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 461 461 N-linked (GlcNAc...) (Potential).  
SQ SEQUENCE 545 AA; 61077 MW; FAC4BDFD85B933 CRC64;

Query Match 59.5%; Score 1624; DB 1; Length 545;  
Best Local Similarity 58.1%; Pred. No. 1.5e-119;  
Matches 311; Conservative 93; Mismatches 115; Indels 16; Gaps 4;  
Qy 3 VLLLLVLL-----LAVPPRTAEQLQGLREPIGAVSPAFSLTLDASLARDPR 50  
:||||| :  
Db 13 LLLLLPLGLGPCSPGTPAAAAPADAAAELEFFTERPLHLVSPAFSLTLDANLATDPR 72  
:||||| :  
Qy 51 FVALLRHPKLHTLASGLSPGFLRFGTSTDPFLIPNPKDSTWEEKV--LSFQAKDVCEA 108  
:||||| :  
Db 73 FFFELGSSKLTARGLAPAYLRFEGNGKDELIFDPKKEPAFEERSYWLQ--SNQDICKS 131  
:||||| :  
Qy 109 WPSFAVVPKLLLTOWPLOEKLLLAHSHKXKNTTITRSTLDILHTFASSGFRVLFGLN 168  
:||||| :  
Db 132 GSIPSDVEEKLRLWEPFQEQVLLREYQOKFTNTSYSSVDMLTYTFASCSGLNIFGVN 191  
:||||| :  
Qy 169 ALLRRAGLQWDSNAKOLLGYCAORSYNI SWELGNPNPSFRKSGICICIDGFGOLGRDFVHL 228  
:||||| :  
Db 192 ALLRTTDMHWDSSNAQLLDYCSSKNYNI SWELGNPNPSFRKAGIFINGQLGEDFIEF 251  
:||||| :  
Qy 229 ROLLSQHPLRYHAELYGLDVGGQPKRHTQHLLRSFPMKSGKKAIDSVTWHYYVNGRSATRE 288  
:||||| :  
Db 252 RKLGLK-SAFKNAKLYGPDIGQPRNTVOMLKSFLKAGGEVIDSVTWHYYVNGRIATKE 310  
:||||| :  
Qy 289 DFLSPVELDSFATAIHVDLGIIVEATVPGKKVWLGTSAYCGGAPOLSNVTVAGFMWLDK 348  
:||||| :  
Db 311 DFLNPDLTDFISSVQKTLRIVEKIRPLKVKWLGTSAYCGGAPFLSNVTVAGFMWLDK 370  
:||||| :  
Qy 349 LGLAARGIDVWROVSFGAGSHLVDAKFKPLDYMLSLLYKRLVGRVLOASVEQADA 408  
:||||| :  
Db 371 LGLSARMGIEVWVKQVLFAGNHYHLVDGNFEPDYMLSLFLKLVGNKVMASVKGPDPR 430  
:||||| :  
Qy 409 RRPVRYLHCTNPRHPKYREGDVTFLALNLSNVTSLOLPKOLWSKSDVOYLLPHGKDSI 468  
:||||| :  
Db 431 SKFRVYLHCTNPKPRYKEGDVTLYALNLSNVTSLOLPKOLWSKSDVOYLLPHGKDSI 490  
:||||| :  
Qy 469 LSRVQLNGRLQLQWVDETPALHEMALAPGSTGLPAFSGYGFVIRNAKAIACI 523  
:||||| :  
Db 491 LSKSVQLNGQILKMWDEQTLPALTEKPLHPGSSGLMPPSPSYGFFVIRNAKAAACI 545  
:||||| :  
RESULT 8  
ID HPSE MOUSE STANDARD; PRT; 535 AA.  
AC Q6YGL; Q8K3K3;  
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
DT 11-OCT-2005, sequence version 2.  
DT 07-MAR-2006, entry version 13.  
DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) [Contains:  
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].  
GN Name=HPse; Synonyms=Hpa;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Murinae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA].  
RC STRAIN=SJL/J; TISSUE=Spleen;  
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;  
RA Hulet M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,  
RA Parish C.R.;  
RT "Cloning of mammalian heparanase, an important enzyme in tumor  
RT invasion and metastasis.";  
RL Nat. Med. 5:803-809(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 28-57 AND 150-179,  
RP GLYCOSYLATION, BIOPHYSICO-CHEMICAL PROPERTIES, ENZYME REGULATION, AND  
RP SUBUNITS.  
RC STRAIN=FVB; TISSUE=Embryo;  
RX MEDLINE=22350326; PubMed=12460766; DOI=10.1016/S1046-5928(02)00558-2;  
RA Mao H.-Q., Navarro E., Patel S., Sargent D., Koo H., Wan H.,  
RA Plata A., Zhou Q., Ludwig D., Bohlen P., Kussie P.;



QY 288 EDFSLSEVLDSFATAIHVDVIGIVEATVPGKKVWLGTSAYGGAPOLSNYYVAGFMWLD 347  
 DB 300 EDFLSSDLVTLFSLVQKILKVTKEITPGKKVWLGTSAYGGAPLSTNTFAAGFMWLD 359  
 QY 348 KGLAARRGIDVVMQVSGFAGSVHLVDAGFKPLPDYWLKLVKRLVGTIVLQASVEQAD 407  
 DB 360 KGLSAQMGIIEVVMQVQFFGAGNHLVDENPEPLDYLWLSLLFKKLVGPRVLLSRVKGPD 419  
 QY 408 ARPRVYLHCTNPRHPKRYRGDVTFLFALNLSNVTQSLQPLKQLMSKSVDOVLLPHGKDS 467  
 DB 420 RSKLRVYLHCTNVYHRYQSGDLTLVNLHNTVKYLKVPPLPKRPVDTYLLKPSGPDG 479  
 QY 468 ILSREVQLNGRLQWVDDELTPALHEMALAPGSTGLPAPFSYGFYVIRNAKAIACI 523  
 DB 480 LLSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSPLPFSYGFYVIRNAKAIACI 535

RESULT 9  
 HPSE\_RAT  
 ID HPSE\_RAT STANDARD; PRT; 536 AA.  
 AC Q71RF1; Q9OZF8;  
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
 DT 05-JUL-2004, sequence version 1.  
 DT 07-MAR-2006, entry version 11.  
 DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) [Contains:  
 DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].  
 GN Name=Hpa; Synonyms=Hep;  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Rattus.  
 ON NCBI\_TaxID=10116;  
 RN [1]\_TaxID=10116;  
 RP NOCLOTIDE SEQUENCE [NRNA].  
 RC TISSUE=Placenta;  
 RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;  
 RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,  
 RA Parish C.R.;  
 RT "Cloning of mammalian heparanase, an important enzyme in tumor  
 RT invasion and metastasis";  
 RL Nat. Med. 5:803-809 (1999).  
 [2]  
 RP NOCLOTIDE SEQUENCE [NRNA], AND ENZYME REGULATION.  
 RX MEDLINE=22194309; PubMed=12077130; DOI=10.1074/jbc.M203282200;  
 RA Podyna-Inoue K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;  
 RT "Characterization of heparanase from a rat parathyroid cell line";  
 RL J. Biol. Chem. 277:32459-32465 (2002).  
 CC -!- FUNCTION: Endoglycosidase which is a cell surface and  
 CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
 CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
 CC proteoglycans. Also implicated in the extravasation of leukocytes  
 CC and tumor cell lines. Contributes to metastasis and angiogenesis  
 CC (By similarity).  
 CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower  
 CC extent, by heparin and sulfamin (By similarity). Activated by  
 CC calcium and magnesium. Inhibited by EDTA.  
 CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
 CC subunits, the proteolytic products (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.  
 CC Secreted, internalised and transferred to late endosomes/lysosomes  
 CC as a proheparanase. In lysosomes, it is processed into the active  
 CC form, the heparanase. The uptake or internalisation of  
 CC proheparanase is mediated by HSPGs. Heparin appears to be a  
 CC competitor and retain proheparanase in the extracellular medium  
 CC (By similarity).  
 CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form  
 CC leads to the generation of a linker peptide, 8 kDa and 50 kDa  
 CC product. The active form, the 8/50 kDa heterodimer, is resistant  
 CC to degradation. Complete removal of the linker peptide appears to  
 CC be a prerequisite to the complete activation of the enzyme (By  
 CC similarity).  
 CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears

CC to be essential for its solubility (By similarity).  
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 CC -----  
 DR EMBL; AF359508; AAQ15189.1; -; mRNA.  
 DR EMBL; AF184967; AAF04563.1; -; mRNA.  
 DR RGD; 61969; Hps.  
 DR InterPro; IPR005199; Glyco\_hydro\_79\_N.  
 DR Pfam; PF03662; Glyco\_hydro\_79n; 1.  
 KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;  
 KW Signal.  
 FT SIGNAL 1 28 By similarity.  
 FT CHAIN 29 102 Heparanase 8 kDa subunit.  
 FT /FTID=PRO\_0000042266.  
 FT PROPEP 103 150 Linker peptide (By similarity).  
 FT /FTID=PRO\_0000042267.  
 FT CHAIN 151 536 Heparanase 50 kDa subunit.  
 FT /FTID=PRO\_0000042268.  
 FT REGION 151 155 Heparin/HS-binding (By similarity).  
 FT REGION 263 273 Heparin/HS-binding (By similarity).  
 FT ACT\_SITE 218 218 Proton donor (Potential).  
 FT ACT\_SITE 336 336 Nucleophile (Potential).  
 FT CARBOHYD 155 155 N-linked (GlcNAc...) (By similarity).  
 FT CARBOHYD 193 193 N-linked (GlcNAc...) (By similarity).  
 FT CARBOHYD 210 210 N-linked (GlcNAc...) (By similarity).  
 FT CARBOHYD 452 452 N-linked (GlcNAc...) (By similarity).  
 FT CONFLICT 15 15 G -> R (in Ref. 2).  
 FT CONFLICT 227 227 H -> Q (in Ref. 2).  
 FT CONFLICT 350 350 D -> N (in Ref. 2).  
 SQ SEQUENCE 536 AA; 60480 MW; C434E04CF536EA4D CRC64;  
 Query Match 57.6%; Score 1572; DB 1; Length 536;  
 Best Local Similarity 58.9%; Pred. No. 2e-115;  
 Matches 302; Conservative 87; Mismatches 118; Indels 6; Gaps 3;  
 QY 14 PPRRTAEQLQLREPIGAVSPAFSLTLDSALDRPRFVALLRHPKHLTASGLSPGFLR 73  
 DB 27 PTKDQVDELYFTKRLFQSVSPSLITDASLATDPRFLTFLGSPRLARALGSLPAYLR 86  
 QY 74 FGGTSTDFLI FPNKOSTWEEKVSEFOAK---DVCSEMPSFVAVPKLLLTQWLPQEKLL 130  
 DB 87 FGGTKTDFLI FDNKPTSEER---SYWQSDNNDICGSRVSADVLRLQWMPFQELL 144  
 QY 131 LAEHSKKKHNVTITRSTDLIHTFASSGFRVLVGLNALLRRAGLOWSSNAKQLGVC 190  
 DB 145 LRSQYQREKFNSTYSRSDVMLYSFAKSRDLILFGLNALLRTPDLRWNSNAQLLLNYC 204  
 QY 191 AQRSYNISWEI GNEPNSFRKSGICIDGFLGRDFVHLRQLLSQHPLYRHAELYGLDVQ 250  
 DB 205 SSKGYNISWEI GNEPNSFWKKAHISIDGLQIGEDFVELHKL--KSAFQNAKLYGPDIG 263  
 QY 251 PRKHTOHLRSFMKSGOKAIDSVTHHYVNGRSATREDFLSPVELDSFATAIHDLVGI 310  
 DB 264 PRGKTVLLRSFLKAGCEVIDSLTWHYLYNGRVATKEDFLSSDVLDTFILLSVQKILVT 323  
 QY 311 EATVPGKKVWLGTSAYGGAPOLSNYYVAGFMWLDKLGAAARRGIDVVMQVSGFAGS 370  
 DB 324 KEMTPGKKVWLGTSAYGGAPLSDTFAAGFMWLDKLGSAQIGIEVVMQVFFGAGN 383  
 QY 371 YHLVDAGFKPLPDYWLKLVKRLVGTIVLQASVEQADARRPRVYLHCTNPRHPKRYRGD 430  
 DB 384 YHLVDENPEPLDYLWLSLLFKKLVGPRVLLSRVKGPDRLVYHCTNVHPRVREGDL 443  
 QY 431 TLFALNLSNVTQSLQPLKQLMSKSVDOVLLPHGKDSILSREVQLNGRLQWVDDELTPA 490  
 DB 444 TLVYLNLHNTVKYLKVPPLPKRPVDTYLLKPSGPDG 503  
 QY 491 LHEMALAPGSTGLPAPFSYGFYVIRNAKAIACI 523  
 DB 504 LTKPLPAGSSLSVPAPFSYGFYVIRNAKAIACI 536









Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
"Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences."  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
[2]

NUCLEOTIDE SEQUENCE.  
RC TISSUE=PCR rescued clones;  
RG NIH MGC Project;  
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; BC112356; AA112357.1; -, mRNA.  
SQ SEQUENCE 592 AA; 66610 MW; 94689E1C2A74359F CRC64;

Query Match 37.7%; Score 1028.5; DB 2; Length 592;  
Best Local Similarity 40.6%; Pred. No. 2.4e-72;  
Matches 232; Conservative 83; Mismatches 199; Indels 57; Gaps 11;  
[2]

1 MLVLLVLLLA-----VPPRTAEQ-----LGLREPIGAVSPAFLSLTLDAS 44  
27 LYLALLHLSSAGDRRPIVDRAAGLKEKTLILLDVSTKNPVRTVNEFLSLQLDPS 86  
45 LARDPRFVALRHPKLHTLASGLSPGFLRFGGTSTDFLIF---NPNKD-----S 90  
87 ITHD-GWLDFLSSKRLVTLARGLSPAFLRFGKRTDFLQFQNLARNPAKSRGCGPDYYLK 145  
91 TWEEKVISEFOAKDV---CEAMPGFVVPKLLLTQWPLQEK-----LLLAHSWKHKH 140  
146 NYEDDIVRSQDVALDKQCK---IAQHPDVMW---ELQREKAAQMHLVLLKEQPSNTYS 198  
141 NTTITTRSTDLTHTFASSSGFRLVFGNLARRAGLOWDSNAKQLLYGCAQRSVNISWE 200  
199 NLIIATSLDKLYNFACDGLHLIFALNALRRPNNSWNSALSLLKYSASKYINISWE 258  
201 LGNEPNSFRKSGGICIDQGLGRDFVHLRQLLSOPLRYHAELVGLDVQGRKHITQHLR 260  
259 LGNEPNRYRTHGKRAVNGSLQCKDYIQLKSLQPIRYSRASLIGPNIQRPKNVIALLD 318  
261 SFMKSGGKAIDSVTWHHYVNGSRATREDFLSPVLDSFATAIHVLGIVEATVPKKVW 320  
319 GFMKVAGSTVDVATVQHCYIDGRVVVKMDFLTKRLDITLSQIRKIKQVVNTYTPGKKIW 378  
321 LGFTGSAYGGAPOLSNITYAGFWMLDKLGAAARGIDVVRQVSFGAGSHVLVDAGFKP 380  
379 LEGVYVTTSGAGTNNLSDSYAAAGFWLNTLGLMANQGDWVTRHSFFDHYGNHVLQDNFP 438  
381 LPDYWLILLYKRLVGTGRVLQASVEQADAR-RP-----RVYLHCTNPRHPKRGEDVT 431  
439 LPDYWLISLLYKRLIGPKVLAVHVLAVGLQRPGRVIRDKLRIYAHTNHHNVTYRGSI 498  
432 LFAINLSNVTOSLQPKQMSKSDVQYLLPHGKDSILSREVQLNGRLLQWDDTETPAL 491  
499 LFIINLHRSRKIKLTGTRDKLVHQYLLQYVQGEGLKSKVQLNGQLVWVDDGTLP 558  
492 HEMALAPGSTLGLPAFSYGFVIVRNAKAIAC 522  
559 KPRPLRAGRITLVIPPTMTGFWKNNVALAC 589

```

RESULT 14
Q4TB80_TETNG
1D Q4TB80_TETNG PRELIMINARY; PRT; 597 AA.
AC Q4TB80;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Chromosome 17 SCAF7180, whole genome shotgun sequence. (Fragment).
GN ORNames=GSTENG0003868001;
OS Neoraoon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Neoteleostei;

```

OC	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC	Tetraodontoidea; Tetraodontidae; Tetraodon.
OX	NCBI_TaxID=99883;
RN	[1]
RP	NUCLEOTIDE SEQUENCE.
RX	PubMed=13495914; DOI=10.1038/nature03025;
RA	Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA	Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA	Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA	Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA	Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA	Bremont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA	Cruaud C., Duprat S., Brotier P., Coutanceau J.-P., Gouzy J.,
RA	Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
RA	Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA	Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA	Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA	Winkler P., Lander E.S., Weissenbach J., Roest Crolihus H.;
RT	"Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT	the early vertebrate proto-karyotype.";
RL	Nature 431:946-957(2004).
RN	[2]
RP	NUCLEOTIDE SEQUENCE.
RG	Genoscope; Whitehead Institute Centre for Genome Research;
RL	Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC	-!- CAUTION: The sequence shown here is derived from an
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC	preliminary data.
CC	-----
CC	Copyrighted by the Uniprot Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a>
CC	Distributed under the Creative Commons Attribution-NoDerivs License
CC	-----
DR	EMBL; CAEA01007180; CAFE98952.1; -; Genomic_DNA.
DR	NON TER 597 597
SQ	SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;

Query Match	34.3%	Score 936.5	DB 2	Length 597	
Best Local Similarity	39.2%	Pred. No. 4.8e-65			
Matches	215	Conservative 80	Mismatches 198	Indels 55	Gaps 11
QY	21	LQLGLEPICANSPAFSLTLDASLARDPRFVALLRHPKLTHTLASHGLSPGLFPGTSTD	80		
DB	58	LDVNTRESPIRVLNDNLSQLDPSIIKD-GWDLFLSSGRLVTLARGLSPAFLFPGKRTD	116		
QY	81	FLIFNPNKO-----STWEKVLVSFOAKDV---CEAWPSFAVVPKLLLTQWP	124		
DB	117	FLOFTQKNLAKRPGCPDYLLKYVEDDIIRSDIALDKQKCK-----LASHPDMML--E	169		
QY	125	LQBK-----LLASHSWKKHKNTTIT-----RSTLDILHTTPASSSGFLVFLGNA	169		
DB	170	LOREKAASTQLVLLKEQLSNIYSNI.TLGLFSHSRIARSLDKLYNFADCAGLHLILGLNA	229		
QY	170	LLRRAGLQWDSSNAKOLLGYCAORSYNIISWELGNENPNSFRKSGICIDGOLFQGRDFVHLR	229		
DB	230	LHRNPDSWNTSSTLSLLKYSACKKYNISWELGNEPNAYRSMVGHVANSLSQAQDYTKLR	289		
QY	230	QLLSQHPLYRHAELYGLDVGOPRKHQTHLLRSFPMKSGGKAIDSVTHWHYYVNGRSATRED	289		
DB	290	TLQSQSVYYSRAQLGYFNAGRPKNALLLDEPFMTKVTGVVDVATVQHTYMDGRIKKVED	349		
QY	290	FLSPYELDSFATAIHDVLGLIVEATVPKKVWLGETSGYGGGAQPLSNTYVAGFMWLDKL	349		
DB	350	FLKTRLLDITLTELQSKVTKVNTHTPGKKVWLGGGLGPATWGTGMSNLSDTFAAGFLVWNTL	409		
QY	350	GLAARGIDVWMR-----QVSGFAGSXLVDAGFKP-LPDYWLISLLYKRLVGTRTVLAQSV	404		
DB	410	GMAAMOGIDVLRQAVQVQHTNKSQVALFLQMFVPSPFDYWFSLVFVKRLVGPVKVLAVRVA	469		
QY	405	QADARPP-----RVYLHCTNPRHPKRYRGDVTFLFALNLSNVTQSLQLPKQLAWSKS	454		
DB	470	GLQ-RKPGQGVTRDKLRIYAHCTSYSNHNYVGRGSTIYIINLHRRSRKKIKLAGTILRNNI	528		
QY	455	VDQYLLPHGKDSILGRESVQLNGRLQLQMWDDETLPALHEMALAPGSGTLGLPAFSGYFYVI	514		

4





GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:31:47 ; Search time 69.589 Seconds

(without alignments)  
65.702 Million cell updates/sec

Title: US-10-645-659A-6

Perfect score: 61

Sequence: 1 CTWTDNPRYK 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

1: A\_Geneseq\_8:\*

2: Geneseqpl980s:\*

3: Geneseqpl990s:\*

4: Geneseqp2000s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

9: Geneseqp2005s:\*

10: Geneseqp2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	100.0	10	8	ADR88212 Human hep
2	61	100.0	10	8	ADT78179 Functiona
3	61	100.0	10	8	AEA42428 Human hep
4	61	100.0	15	9	ADU71246 Human hep
5	61	100.0	15	9	ADU70983 Human hep
6	61	100.0	15	9	ADU71087 Human hep
7	61	100.0	15	9	ADU71245 Human hep
8	61	100.0	386	8	ADR88207 Human mat
9	61	100.0	386	8	ADT78174 45kDa sub
10	61	100.0	386	9	ADY27057 Heparanas
11	61	100.0	386	9	ADZ18995 Human hep
12	61	100.0	386	9	AEA42423 Human mat
13	61	100.0	460	9	ADY27061 Heparanas
14	61	100.0	486	9	AE887589 Human hep
15	61	100.0	492	9	ADZ18996 Hep106 co
16	61	100.0	493	9	AE887562 Human hep
17	61	100.0	495	9	ADZ18999 Hep109 co
18	61	100.0	497	9	AE887587 Human hep
19	61	100.0	501	9	ADZ19000 HepGS3 co
20	61	100.0	507	9	ADZ19005 HepGS6 co
21	61	100.0	508	9	ADY27058 Human ina
22	61	100.0	526	9	ADZ19006 Hepiyalur
23	61	100.0	527	5	ABB07815 Chicken s

24	61	100.0	527	7	ABW02018
25	61	100.0	527	8	ADO63825
26	61	100.0	527	8	ADO63827
27	61	100.0	527	8	ADO63826
28	61	100.0	527	9	ADZ19004
29	61	100.0	530	2	AAZ34173
30	61	100.0	532	2	AAZ17083
31	61	100.0	543	2	AAZ02345
32	61	100.0	543	2	AAZ17082
33	61	100.0	543	3	AAZ57590
34	61	100.0	543	3	AAZ08849
35	61	100.0	543	3	AAZ52990
36	61	100.0	543	4	AAZ97635
37	61	100.0	543	4	AAZ86206
38	61	100.0	543	4	AAZ88361
39	61	100.0	543	5	ABB07813
40	61	100.0	543	7	ADD18950
41	61	100.0	543	7	ADG88800
42	61	100.0	543	8	ADL16379
43	61	100.0	543	8	ADK52086
44	61	100.0	543	8	ADM48716
45	61	100.0	543	8	ADM48759
46	61	100.0	543	8	ADN05074
47	61	100.0	543	8	ADN04902
48	61	100.0	543	8	ADO63831
49	61	100.0	543	8	ADO63824
50	61	100.0	543	8	ADO63823
51	61	100.0	543	8	ADO63832
52	61	100.0	543	8	ADO63822
53	61	100.0	543	8	ADQ80372
54	61	100.0	543	8	ADR88210
55	61	100.0	543	8	ADP25079
56	61	100.0	543	8	ADT78177
57	61	100.0	543	9	ADY27036
58	61	100.0	543	9	ADY63087
59	61	100.0	543	9	AEA42466
60	61	100.0	543	9	AEA42426
61	61	100.0	543	10	AE966848
62	61	100.0	545	6	ABP56822
63	61	100.0	545	7	ADE16012
64	61	100.0	545	8	ADL93951
65	61	100.0	556	9	ADZ19010
66	61	100.0	570	9	ADZ19008
67	61	100.0	588	2	AAZ30124
68	61	100.0	592	2	AAZ02346
69	61	100.0	592	3	AAZ08850
70	61	100.0	592	7	ADG88804
71	61	100.0	592	8	ADL16383
72	61	100.0	592	8	ADM48720
73	61	100.0	592	9	AEA42461
74	56	91.8	9	9	ADU70671
75	56	91.8	15	9	ADU70848
76	49	80.3	15	9	ADU70982
77	49	80.3	545	9	ADY27034
78	47	77.0	9	9	ADU70619
79	47	77.0	15	9	ADU71247
80	44	72.1	9	9	ADU70744
81	44	72.1	15	9	ADU71086
82	44	72.1	819	8	ADN24558
83	44	72.1	820	8	ADN21798
84	43	70.5	487	8	ADN25773
85	43	70.5	639	3	AAZ74439
86	43	70.5	639	3	AAZ74438
87	43	70.5	639	3	AAZ74440
88	43	70.5	648	6	ABP80492
89	43	70.5	820	8	ADZ27119
90	43	70.5	829	8	ADZ26735
91	43	70.5	839	8	ADN26572
92	43	70.5	839	8	ADN26757
93	43	70.5	858	8	ADZ24760
94	43	70.5	860	9	ADU81528
95	43	70.5	860	9	ADU81526
96	43	70.5	864	8	ADZ26367

Abw02018	Chimeric
Ado63825	Chimeric
Ado63827	Chimeric
Ado63826	Chimeric
Adz19004	HepGS4 co
Aay34173	Human pre
Aay17083	Seq ID No
Aay02345	A human h
Aay17082	Human hep
Aay57590	Human hep
Aab08849	Amino aci
Aay52990	Human hep
Aay97635	Human hep
Aab86206	Human hep
Aab88361	Human mem
Abb07813	Human hep
Add18950	Human dis
Adg88800	Human hpa
Adl16379	Human hep
Adk52086	Human ato
Adm48716	Human hpa
Adm48759	Human hpa
Adn05074	Antipsori
Adn04902	Antipsori
Ado63831	Human hep
Ado63824	Human hep
Ado63823	Human hep
Ado63832	Human hep
Ado63822	Human hep
Adq80372	Heparanas
Adr88210	Human pre
Adp25079	PRO polyp
Adt78177	Human hep
Ady27036	Human hep
Ady63087	Human clo
Aea42466	Human hep
Aea42426	Human hep
Aee966848	Human hep
Abp56822	Human hep
Adel6012	G-coupled
Adl93951	Human G-c
Adz19010	Heparanas
Adz19008	Heparanas
Aay30124	A human p
Aay02346	A human h
Aab08850	Amino aci
Adg88804	Human SK-
Adl16383	Human hep
Adm48720	Human SK-
Aea42461	Human hep
Adu70671	Human hep
Adu70848	Human hep
Adu70982	Human hep
Ady27034	Bovine he
Adu70619	Human hep
Adu71247	Human hep
Adu70744	Human hep
Adu71086	Human hep
Adn24558	Bacterial
Adn21798	Bacterial
Adn25773	Bacterial
Aay74439	Neisseria
Aay74438	Neisseria
Aay74440	Neisseria
Abp80492	N. gonorr
Adz27119	Bacterial
Adz26735	Bacterial
Adn26572	Bacterial
Adn26757	Bacterial
Adz24760	Bacterial
Adu81528	Aconitase
Adu81526	Aconitase
Adz26367	Bacterial

97	43	70.5	865	9	ADU81525	Adu81525 Aconitase	170	38	62.3	871	8	ADS24438	Bacterial
98	43	70.5	866	9	ADU81527	Adu81527 Aconitase	171	38	62.3	882	8	ADN25341	Bacterial
99	43	70.5	867	9	ADU81524	Adu81524 Aconitase	172	38	62.3	884	8	ADS21824	Bacterial
100	43	70.5	868	3	AY93301	Amino aci	173	38	62.3	887	8	ADN24259	Bacterial
101	43	70.5	868	3	ADU81522	Adu81522 Aconitase	174	38	62.3	887	8	ADN25875	Bacterial
102	43	70.5	868	3	ADU81523	Adu81523 Aconitase	175	38	62.3	887	8	ADN28752	Bacterial
103	43	70.5	869	3	AY93285	Amino aci	176	38	62.3	887	8	ADY24031	Plant ful
104	43	70.5	871	7	ABO80676	Pseudomon	177	38	62.3	889	5	ABB57134	Mouse isc
105	43	70.5	875	6	ADA33597	Acinetoba	178	38	62.3	889	8	ADN04279	Antipsori
106	41	67.2	9	9	ADU70745	Human hep	179	38	62.3	889	8	ADP24012	PRO polyv
107	41	67.2	15	5	ADU71294	Human hep	180	38	62.3	891	2	AAR84338	Maize aco
108	40	65.6	297	5	ABBS4966	Lactococc	181	38	62.3	891	4	AAU29353	Novel mar
109	40	65.6	297	8	ADS29432	Bacterial	182	38	62.3	891	4	ADS42664	Bacterial
110	40	65.6	532	4	AAG92514	C glutami	183	38	62.3	891	8	ADN18158	Bacterial
111	40	65.6	881	5	ABP28198	Streptoco	184	38	62.3	891	9	AEB40161	L. pneumo
112	40	65.6	881	8	ADV89585	Streptoco	185	38	62.3	891	9	AEB36786	L. pneumo
113	40	65.6	881	8	ADV82992	Streptoco	186	38	62.3	892	6	ABM67426	Phototrab
114	40	65.6	881	8	ADV80838	Streptoco	187	38	62.3	893	8	ADN22413	Bacterial
115	39	63.9	960	6	ABU22866	Protein e	188	38	62.3	895	7	ABO61376	Klebsiell
116	39	63.9	299	9	AEC21206	L. acidop	189	38	62.3	896	8	ADS23030	Bacterial
117	39	63.9	301	9	AEC21150	L. acidop	190	38	62.3	896	8	ADN17556	Bacterial
118	39	63.9	378	7	ABO63658	Klebsiell	191	38	62.3	896	8	ADN21724	Bacterial
119	39	63.9	380	2	AY17085	Rat hepar	192	38	62.3	896	8	ADN24478	Bacterial
120	39	63.9	536	5	ABO07812	Rat hepar	193	38	62.3	897	8	ADF54681	Aconitase
121	39	63.9	536	8	ADR88209	Rat hepar	194	38	62.3	898	8	ADQ26200	Arabidops
122	39	63.9	536	8	ADY78176	Rat hepar	195	38	62.3	899	4	ABB62277	Drosophil
123	39	63.9	536	9	ADY27035	Rat hepar	196	38	62.3	899	5	ABP66115	Bifidobac
124	39	63.9	536	9	AEA42425	Rat hepar	197	38	62.3	900	5	ABB48185	Listeria
125	39	63.9	2931	4	ABB68229	Drosophil	198	38	62.3	901	4	AAG81624	S. epider
126	38	62.3	74	3	AAG01360	Human sec	199	38	62.3	901	4	AAB83182	Coryneb
127	38	62.3	90	4	AAM25264	Human pro	200	38	62.3	901	6	ABM72914	Staphyloc
128	38	62.3	256	4	ABB64501	Drosophil	201	38	62.3	901	8	ADS42317	Bacterial
129	38	62.3	380	2	AY17084	Mouse hep	202	38	62.3	901	8	ADN26748	Bacterial
130	38	62.3	384	5	ABP26811	Streptoco	203	38	62.3	901	8	ADN26679	Bacterial
131	38	62.3	384	8	ADV88250	Streptoco	204	38	62.3	902	4	ABB61080	Drosophil
132	38	62.3	384	8	ADY79503	Streptoco	205	38	62.3	903	6	ABR64210	Angiogene
133	38	62.3	384	8	ADV81674	Streptoco	206	38	62.3	903	8	ADR28047	Bacterial
134	38	62.3	462	8	ADX88784	Plant ful	207	38	62.3	903	8	ADU06553	Novel bro
135	38	62.3	487	8	ADY73911	Plant ful	208	38	62.3	903	9	ADY70686	Human bac
136	38	62.3	517	6	ADA89649	Staphyloc	209	38	62.3	904	7	ADF04200	Bacterial
137	38	62.3	535	3	AAB08851	A murine	210	38	62.3	904	8	ADF54679	Aconitase
138	38	62.3	535	5	ABO07811	Mouse hep	211	38	62.3	906	8	ADN17965	Bacterial
139	38	62.3	535	7	ADG88834	Mouse hpa	212	38	62.3	906	8	ADY24027	Plant ful
140	38	62.3	535	8	ADL16413	Mouse hep	213	38	62.3	908	8	ADY28475	Bacterial
141	38	62.3	535	8	ADM48750	Mouse hpa	214	38	62.3	909	8	ADO59731	B. subtil
142	38	62.3	535	8	ADR88208	Mouse hep	215	38	62.3	909	8	ADS44720	Bacterial
143	38	62.3	535	8	ADY78175	Mouse hep	216	38	62.3	911	8	ADS24683	Bacterial
144	38	62.3	535	9	ADY27033	Murine he	217	38	62.3	914	5	ABP40363	Staphyloc
145	38	62.3	535	9	AEA42424	Mouse hep	218	38	62.3	914	8	ADS04961	Staphyloc
146	38	62.3	553	4	AAB79520	Coryneb	219	38	62.3	914	8	ADS26445	Bacterial
147	38	62.3	557	4	AAB79519	Coryneb	220	38	62.3	915	8	ADN25706	Bacterial
148	38	62.3	573	8	ADY71355	Plant ful	221	38	62.3	915	8	ADS26815	Bacterial
149	38	62.3	631	8	ADY73304	Plant ful	222	38	62.3	919	2	AAR84337	Arabidops
150	38	62.3	643	8	ADY74277	Plant ful	223	38	62.3	922	2	ADS27347	Bacterial
151	38	62.3	655	8	ADY71025	Plant ful	224	38	62.3	927	6	AAR85598	Arabidops
152	38	62.3	668	5	AAM49526	B. mori t	225	38	62.3	927	6	ADA36635	Acinetoba
153	38	62.3	680	8	ADY77402	Plant ful	226	38	62.3	929	8	ADS22231	Bacterial
154	38	62.3	732	3	AAB36178	Mouse sev	227	38	62.3	934	9	ADY60907	Abiotic s
155	38	62.3	732	8	ADO28908	Mouse nov	228	38	62.3	941	8	ADS30597	Bacterial
156	38	62.3	797	6	ABM65102	Propionib	229	38	62.3	943	8	AAG91445	C glutami
157	38	62.3	805	4	AAU62827	Propionib	230	38	62.3	943	7	ADD13587	C. glutam
158	38	62.3	805	6	ABM59346	Propionib	231	38	62.3	943	9	AEB13175	C. glutam
159	38	62.3	824	8	ADY75932	Plant ful	232	38	62.3	968	8	ADX75970	Plant ful
160	38	62.3	848	5	ABBS3982	Lactococc	233	38	62.3	981	7	ABO68052	Pseudomon
161	38	62.3	848	8	ADS29320	Bacterial	234	38	62.3	1053	6	ABM70125	Phototrab
162	38	62.3	852	2	AYY30948	Human E3	235	38	62.3	1298	9	ADY61065	Abiotic s
163	38	62.3	854	2	AAV30949	Murine E3	236	37	60.7	9	ADU70496	Human hep	
164	38	62.3	854	9	AEA18809	Amino aci	237	37	60.7	15	ADU70883	Human hep	
165	38	62.3	855	9	AED00545	Sulfolobu	238	37	60.7	74	4	ABG25267	Novel hum
166	38	62.3	858	9	ADN26062	Bacterial	239	37	60.7	75	6	ABO00942	Polypepti
167	38	62.3	862	9	AEA18808	Amino aci	240	37	60.7	84	7	ABO67476	Klebsiell
168	38	62.3	869	8	ADS29065	Bacterial	241	37	60.7	97	4	AAU39606	Propionib
169	38	62.3	870	8	ADS44479	Bacterial	242	37	60.7	97	6	ABM36125	Propionib

243	37	60.7	109	2	AAW93582	Aaw93582 Rat rAPO4	316	35	57.4	297	3	AAG56430	Aag56430 Arabidops
244	37	60.7	197	6	ADA35040	Ada35040 Acinetoba	317	35	57.4	317	3	AAG60489	Aag60489 Arabidops
245	37	60.7	198	8	ADX67422	Adx67422 Plant ful	318	35	57.4	317	3	AAG56429	Aag56429 Arabidops
246	37	60.7	384	2	AAI24914	Aay24914 Eibenia f	319	35	57.4	354	3	ABF42527	Abf42527 Human ORF
247	37	60.7	730	3	AAAB36179	Aab36179 Rat seven	320	35	57.4	423	5	ABP73924	Abp73924 Candida a
248	37	60.7	756	6	AAE330471	Aae330471 Haemophil	321	35	57.4	423	7	ABD70164	Abd70164 C. neofo
249	37	60.7	784	7	AAE331168	Aae331168 Human dia	322	35	57.4	428	6	ABR53385	AbR53385 Protein s
250	37	60.7	805	8	ADN23021	Adn23021 Bacterial	323	35	57.4	428	7	ADK63326	Adk63326 Diseasee t
251	37	60.7	805	8	ADN23020	Adn23020 Bacterial	324	35	57.4	487	3	AAG28042	Aag28042 Arabidops
252	37	60.7	1192	4	ABG02038	Abg02038 Novel hum	325	35	57.4	540	5	AAM49664	Aam49664 Rat SmGlu
253	37	60.7	1653	7	ADF74137	Adf74137 Human nov	326	35	57.4	554	3	ABAB20994	Abab20994 Human Rec
254	37	60.7	2113	4	ABBE64885	Abb64885 Drosophil	327	35	57.4	564	2	AAW48788	Aaw48788 Thyroid p
255	37	60.7	2113	6	ADS963194	Ads963194 Drosophil	328	35	57.4	569	3	AAG28041	Aag28041 Arabidops
256	36.5	59.8	143	6	ABR42559	AbR42559 Novobioci	329	35	57.4	569	3	AAG38775	Aag38775 Arabidops
257	36	59.0	9	9	ADU70820	Adu70820 Human hep	330	35	57.4	621	3	AAG28040	Aag28040 Arabidops
258	36	59.0	15	9	ADU71088	Adu71088 Human hep	331	35	57.4	623	3	AAG38774	Aag38774 Arabidops
259	36	59.0	202	8	ADQ08768	Adq08768 Ciona int	332	35	57.4	624	2	AAW48789	Aaw48789 Thyroid p
260	36	59.0	218	8	ADH39781	Adh39781 Streptomy	333	35	57.4	654	7	ADC31447	Adc31447 Human nov
261	36	59.0	252	3	AAG28249	Aag28249 Arabidops	334	35	57.4	681	2	AAW48786	Aaw48786 Thyroid p
262	36	59.0	252	3	AKG46618	Akg46618 Arabidops	335	35	57.4	689	2	AAW48787	Aaw48787 Thyroid p
263	36	59.0	262	3	AAG46617	Aag46617 Arabidops	336	35	57.4	689	2	ADS23685	Ads23685 Bacteri
264	36	59.0	262	3	AAG28248	Aag28248 Arabidops	337	35	57.4	701	6	ADA34430	Ada34430 Acinetoba
265	36	59.0	279	7	ADF04263	Adf04263 Bacterial	338	35	57.4	731	3	AAG38773	Aag38773 Arabidops
266	36	59.0	341	3	AKG46616	Akg46616 Arabidops	339	35	57.4	740	2	AAW48790	Aaw48790 Thyroid p
267	36	59.0	341	3	AKG28247	Akg28247 Arabidops	340	35	57.4	752	8	ADX95153	Adx95153 Plant ful
268	36	59.0	341	5	ABB93596	Abb93596 Herbicida	341	35	57.4	756	3	AAAB20995	Aab20995 Human Rec
269	36	59.0	463	3	AAV58263	Aav58263 Guar gala	342	35	57.4	784	2	AAW48783	Aaw48783 Thyroid p
270	36	59.0	470	3	AAAB05940	Aab05940 Protein d	343	35	57.4	785	8	ABM83658	Abm83658 Human dia
271	36	59.0	481	2	AAW72095	Aaw72095 Human mGL	344	35	57.4	813	6	ABOS2993	Abos2993 Human put
272	36	59.0	523	5	ABB07814	Abb07814 Chicken h	345	35	57.4	842	8	ABM83657	Abm83657 Human dia
273	36	59.0	523	7	ABW02017	Abw02017 Chicken h	346	35	57.4	848	2	AAW07733	Aay07733 Human thy
274	36	59.0	523	8	ADR88211	Adr88211 Chicken h	347	35	57.4	852	2	AAW48782	Aaw48782 Thyroid p
275	36	59.0	523	8	ADY78178	Ady78178 Chicken h	348	35	57.4	868	4	AAU35640	Aau35640 Haemophil
276	36	59.0	523	9	ADY27037	Ady27037 Chicken h	349	35	57.4	868	6	ABU30578	Abu30578 Protein e
277	36	59.0	523	9	AEA42427	Aea42427 Chicken h	350	35	57.4	881	2	AAW48791	Aaw48791 Thyroid p
278	36	59.0	577	8	ADM48198	Adm48198 Polypepti	351	35	57.4	888	7	ADJ47732	Ade47732 Human NOV
279	36	59.0	594	8	ADO07212	Ado07212 Streptoco	352	35	57.4	888	8	ADJ79002	Adj79002 Human NOV
280	36	59.0	657	8	ADO07196	Ado07196 Streptoco	353	35	57.4	915	2	AAAR80479	Aar80479 Rat metab
281	36	59.0	659	7	ADF04463	Adf04463 Bacterial	354	35	57.4	915	6	AAE30199	Aae30199 Rat metab
282	36	59.0	728	4	AAAG98830	Aag98830 E. coli g	355	35	57.4	915	7	ADD48384	Add48384 Rat Prote
283	36	59.0	728	6	ABU14683	Abu14683 Protein e	356	35	57.4	915	7	ADE55967	Ade55967 Rat Prote
284	36	59.0	867	2	AAW72093	Aaw72093 Human mGL	357	35	57.4	924	2	AAO29104	Ado29104 Mouse nov
285	36	59.0	898	9	ADZ67613	Adz67613 Chimeric	358	35	57.4	924	2	AAAR35445	Aar35445 Human TPO
286	36	59.0	907	9	ADZ67611	Adz67611 Chimeric	359	35	57.4	933	2	AAAR44615	Aar44615 Human thy
287	36	59.0	910	9	ADZ67609	Adz67609 Chimeric	360	35	57.4	933	2	AAAR32875	Aar32875 Human TPO
288	36	59.0	912	9	ADZ67607	Adz67607 Chimeric	361	35	57.4	933	2	AAAR75689	Aar75689 Human thr
289	36	59.0	913	10	AEF72174	Aee72174 Human tar	362	35	57.4	933	7	ADJ68762	Adj68762 Human hea
290	36	59.0	915	2	AAAR72097	Aar72097 Human mGL	363	35	57.4	933	8	ADQ14321	Adq14321 Human thy
291	36	59.0	915	5	ABG95165	Abg95165 Human GPC	364	35	57.4	933	8	ADR41768	Adr41768 Human thy
292	36	59.0	915	5	ABG95155	Abg95155 Human GPC	365	35	57.4	933	8	ADR14361	Adr14361 Human thy
293	36	59.0	915	5	ABG95164	Abg95164 Human GPC	366	35	57.4	933	9	ABE77781	Aeb77781 Human thy
294	36	59.0	915	5	ABG95166	Abg95166 Human GPC	367	35	57.4	933	10	AEQ07493	Aeg07493 Human thy
295	36	59.0	915	5	ABG95163	Abg95163 Human GPC	368	35	57.4	948	2	AAW48781	Aaw48781 Thyroid p
296	36	59.0	915	6	ABP81849	Abp81849 Human met	369	35	57.4	1142	8	ABM84821	Abm84821 Human dia
297	36	59.0	915	7	ADE55969	Ade55969 Human Pro	370	35	57.4	1144	7	ADD01202	Add01202 Human nuc
298	36	59.0	915	7	ADJ93191	Adj93191 Human met	371	35	57.4	1208	2	AAW95050	Aaw95050 Human hel
299	36	59.0	915	8	ADO23103	Ado23103 Human nov	372	35	57.4	1208	3	AAAB20993	Aab20993 Human Rec
300	36	59.0	915	9	ADZ67603	Adz67603 Human met	373	35	57.4	1208	5	ABG93387	Abg93387 Human Rec
301	36	59.0	915	9	AEAB7480	Aeb7480 Human met	374	35	57.4	1208	8	ADQ21562	Adq21562 Human sof
302	36	59.0	915	10	AEET21172	Aee721172 Human tar	375	35	57.4	1208	8	ADU80701	Adu80701 Human RTS
303	36	59.0	922	2	AAAR72098	Aar72098 Human mGL	376	35	57.4	1215	6	ABM15836	Abm15836 Mycobacte
304	36	59.0	922	10	AEET21173	Aee721173 Human tar	377	35	57.4	1291	4	ABG28604	Abg28604 Novel hum
305	36	59.0	1028	7	ADF28592	Adf28592 Murine PA	378	35	57.4	1306	8	ADM90956	Adm90956 Human pha
306	35	57.4	11	2	AAV17080	Aav17080 Human pla	379	35	57.4	3913	6	ABM67350	Abm67350 Photorhab
307	35	57.4	68	6	AAU63347	Aau63347 Propionib	380	35	57.4	10182	5	ABP38314	Abp38314 Scaphyloc
308	35	57.4	68	6	ABM59866	Abm59866 Propionib	381	34	55.7	17	9	ADW26216	Adw26216 Human EA4
309	35	57.4	88	9	AEC37224	Aec37224 ReIE-toxi	382	34	55.7	17	9	AEA17140	Aea17140 Human TNF
310	35	57.4	120	5	ABP02418	Abp02418 Human ORF	383	34	55.7	17	9	ABE17203	Aeb17203 Epha2-spe
311	35	57.4	185	3	AAG56431	Aag56431 Arabidops	384	34	55.7	17	9	AEA43047	Aea43047 Epha2 ant
312	35	57.4	185	3	AAAG60491	Aag60491 Arabidops	385	34	55.7	19	10	AEF39151	Aef39151 Human ser
313	35	57.4	213	5	ABP07167	Abp07167 Human ORF	386	34	55.7	22	10	AEF06087	Aef06087 Human hea
314	35	57.4	285	8	ADN47107	Adn47107 Thermococ	387	34	55.7	71	5	AAU98096	Aau98096 Human bet
315	35	57.4	297	3	AAAG60490	Aag60490 Arabidops	388	34	55.7	78	7	ADH86762	Adh86762 Enterococ

389	34	55.7	98	4	AAB80990	Human ant	462	34	55.7	937	4	AAU02144	Aau02144 Rx 28, mo
390	34	55.7	98	5	ABG78169	Human Fv	463	34	55.7	937	4	AAU02145	Aau02145 Rx 72, mo
391	34	55.7	98	5	ABG78168	Human Fv	464	34	55.7	937	4	AAU02147	Aau02147 Rx 193, m
392	34	55.7	98	5	ABG91860	Human ant	465	34	55.7	937	4	AAU02146	Aau02146 Rx 39, mo
393	34	55.7	98	5	ABG91859	Human ant	466	34	55.7	937	4	AAU02143	Aau02143 Rx 25, mo
394	34	55.7	98	6	ABO27114	Human ger	467	34	55.7	937	4	AAU02148	Aau02148 Rx 7, mod
395	34	55.7	98	7	ADF10159	Antibody	468	34	55.7	938	3	AAU45004	Protein e
396	34	55.7	98	7	ADF09951	Antibody	469	34	55.7	1402	8	ADN22849	Bacterial
397	34	55.7	98	7	ADF10056	VEGF anti	470	34	55.7	2539	8	ADU25430	Bacterial
398	34	55.7	98	7	ADJ80331	VH gene 1	471	33.5	54.9	674	4	ABB61427	L. acidop
399	34	55.7	98	9	ADY75336	Protein e	472	33.5	54.9	1002	3	AAU79166	Pneumocys
400	34	55.7	98	9	AEb13612	Human var	473	33	54.1	44	5	AAE21953	Human lmm
401	34	55.7	98	9	AEC30851	Human var	474	33	54.1	49	2	AAW28322	Staphyloc
402	34	55.7	98	9	AEC36757	Human ger	475	33	54.1	53	5	ABP04205	Human ORF
403	34	55.7	98	10	AE847570	Human CDR	476	33	54.1	55	4	ABBS0601	Human sec
404	34	55.7	100	10	AEF81801	Human ant	477	33	54.1	55	6	ABO44858	Novel hum
405	34	55.7	109	9	ADU26551	Human ant	478	33	54.1	55	7	ABO26338	Protein a
406	34	55.7	117	9	AEa17137	Human TNF	479	33	54.1	57	5	ABP08149	Human ORF
407	34	55.7	119	7	ABR61523	Human RF-	480	33	54.1	63	4	ABU64687	Propionib
408	34	55.7	120	2	AAR65161	Human bcl	481	33	54.1	63	6	ABM61206	Propionib
409	34	55.7	120	7	ABR83200	Human ant	482	33	54.1	92	3	AAU91994	Mutant hu
410	34	55.7	123	9	ADW26196	Human EA4	483	33	54.1	93	4	AAO10801	Human pol
411	34	55.7	123	9	AEb17201	EphA2-spe	484	33	54.1	100	4	ABBS0599	Human sec
412	34	55.7	123	9	AEa43045	EphA2 ant	485	33	54.1	100	6	ABO44856	Novel hum
413	34	55.7	126	2	AAU21147	Human bcl	486	33	54.1	100	7	ABO26336	Protein a
414	34	55.7	136	5	ABB92774	Herbicida	487	33	54.1	102	4	AAO12801	Human pol
415	34	55.7	146	4	AAb20318	Wheat apo	488	33	54.1	124	4	ABB60176	Drosophil
416	34	55.7	149	6	ABJ36935	Anti-CD40	489	33	54.1	130	8	ADX75159	Plant ful
417	34	55.7	152	9	AEC81264	Human mon	490	33	54.1	132	3	AAU62616	Propionib
418	34	55.7	160	7	ADE09227	Novel pro	491	33	54.1	133	4	AAU62616	Propionib
419	34	55.7	200	9	AEb17262	EphA4-spe	492	33	54.1	133	6	ABMS9135	Propionib
420	34	55.7	233	7	ADF07924	Bacterial	493	33	54.1	164	3	AAU30246	Arabidops
421	34	55.7	238	9	ADW26212	Human scF	494	33	54.1	178	8	ADY05698	Plant ful
422	34	55.7	239	2	AAW49692	Human scF	495	33	54.1	181	10	AEF65943	B. subtil
423	34	55.7	259	9	ADW26210	Human scF	496	33	54.1	185	4	AAU35490	Haemophil
424	34	55.7	260	6	ABU19436	Protein e	497	33	54.1	185	6	ABU30345	Protein e
425	34	55.7	261	6	ABU21635	Protein e	498	33	54.1	201	4	ABBS6567	Drosophil
426	34	55.7	262	6	ABU22621	Protein e	499	33	54.1	202	3	AAU30245	Arabidops
427	34	55.7	274	6	ABM70789	Staphyloc	500	33	54.1	209	8	ADX73581	Plant ful
428	34	55.7	279	9	ADW26208	Human scF	501	33	54.1	213	7	ABM73698	DNA clone
429	34	55.7	288	5	ABBS3498	Lactococc	502	33	54.1	213	8	ADY05303	Plant ful
430	34	55.7	289	8	ADS29274	Bacterial	503	33	54.1	224	4	ABM67474	Amino aci
431	34	55.7	359	9	AEa43058	Single ch	504	33	54.1	224	8	ADU04688	Mumps vir
432	34	55.7	373	6	ABM70683	Photorhab	505	33	54.1	225	8	ADT87034	Yeast Str
433	34	55.7	396	4	AAU02152	Plant res	506	33	54.1	263	8	ADX78221	Plant ful
434	34	55.7	430	4	ABBS6838	Drosophil	507	33	54.1	263	8	ADY07670	Plant ful
435	34	55.7	441	4	ABG02026	Novel hum	508	33	54.1	269	9	AEBA2538	L. pneumo
436	34	55.7	449	3	AAU22958	Arabidops	509	33	54.1	275	6	ABU94235	Trifolium
437	34	55.7	455	4	ABB71816	Drosophil	510	33	54.1	276	2	AAU37171	Amino aci
438	34	55.7	467	3	AAU40087	Arabidops	511	33	54.1	306	3	AAU09000	Arabidops
439	34	55.7	467	3	AAU41408	Arabidops	512	33	54.1	306	3	AAU42634	Arabidops
440	34	55.7	467	3	AAU22957	Arabidops	513	33	54.1	325	2	AAW97705	Staphyloc
441	34	55.7	485	2	AAW01459	Arabidops	514	33	54.1	330	3	AAU16019	E. coli p
442	34	55.7	485	3	AAU40086	Arabidops	515	33	54.1	330	4	AAU98321	Escherich
443	34	55.7	485	3	AAU41407	Arabidops	516	33	54.1	334	4	AAU19853	Human nov
444	34	55.7	485	7	ADU75199	Thale cre	517	33	54.1	334	4	AAU87678	Human nov
445	34	55.7	485	8	ADN72983	Thale cre	518	33	54.1	334	5	ABP48073	Human pol
446	34	55.7	489	3	AAU22956	Arabidops	519	33	54.1	334	7	ADU11035	Human pro
447	34	55.7	497	3	AAU41406	Arabidops	520	33	54.1	334	8	ADU15493	Novel hum
448	34	55.7	497	7	ADF06409	Bacterial	521	33	54.1	342	8	ADP04517	Sea squir
449	34	55.7	508	3	AAU40085	Arabidops	522	33	54.1	363	3	AAU77125	Human neu
450	34	55.7	510	8	ADN21556	Bacterial	523	33	54.1	387	4	AAU94592	Human pro
451	34	55.7	578	2	AAU81844	Human afa	524	33	54.1	388	7	ABO74233	Pseudomon
452	34	55.7	599	2	AAU81845	Human afa	525	33	54.1	389	7	ADM05762	Human pro
453	34	55.7	599	2	AAU75655	Human afa	526	33	54.1	389	7	AEC88692	Human CDR
454	34	55.7	634	6	ABM67337	Photorhab	527	33	54.1	391	6	ABU94276	Trifolium
455	34	55.7	643	8	ADL91024	Rat manno	528	33	54.1	391	8	ADU03908	White clo
456	34	55.7	669	6	ABU44839	Protein e	529	33	54.1	399	7	ADF05840	Bacterial
457	34	55.7	679	9	AED53582	Trypanoso	530	33	54.1	410	8	ADR10318	Human pro
458	34	55.7	736	5	ABP65381	Bifidobac	531	33	54.1	423	2	AAU91305	Transcrip
459	34	55.7	751	8	ADK16547	Nanoarcha	532	33	54.1	423	8	ADQ29657	Human col
460	34	55.7	912	3	AAU44818	Potato Gp	533	33	54.1	423	8	ADP24748	PRO polyP
461	34	55.7	937	2	AAU52152	Potato re	534	33	54.1	423	9	AEA04485	Human pro

535	33	54.1	423	10	AEF69969	Aef69969 Coloretta	608	33	54.1	3144	2	AAW36887	Aaw36887 Previousel
536	33	54.1	428	5	ABB90572	Abb90572 Chlamydia	609	33	54.1	3144	2	AAW09871	Aaw09871 Human hun
537	33	54.1	428	6	ABU27068	Abu27068 Protein e	610	33	54.1	3144	2	AAW44742	Aaw44742 Human hun
538	33	54.1	428	9	AEC95688	Aec95688 C. pneumo	611	33	54.1	3144	2	AYX33493	Aay33493 Human hun
539	33	54.1	452	2	ACY35435	Acy35435 C. pneumo	612	33	54.1	3144	9	ADY98141	Ady98141 Human hun
540	33	54.1	453	8	ADR95252	Adr95252 Novel S.	613	33	54.1	3144	9	AEC36191	Aec36191 Amino aci
541	33	54.1	453	9	AEA59122	Aea59122 Streptoco	614	33	54.1	3223	4	ABB11407	Abb11407 Human hun
542	33	54.1	460	7	ADF06963	Adf06963 Bacterial	615	33	54.1	3223	4	ABB11470	Abb11470 Human hun
543	33	54.1	482	2	ARF75382	Aar75382 Natural r	616	33	54.1	3572	5	ABG95659	Abg95659 Human nuc
544	33	54.1	482	2	ARF75386	Aar75386 Natural r	617	33	54.1	4970	9	ABG95659	Abg95659 Human nuc
545	33	54.1	488	8	ADT59749	Adt59749 Plant pol	618	32.5	53.3	537	7	ABO81643	Abob81643 Pseudomon
546	33	54.1	500	4	AAU19735	Aau19735 Human nov	619	32.5	53.3	2359	6	ABP56959	Abp56959 E. maxima
547	33	54.1	500	5	ABP47955	Abp47955 Human pol	620	32.5	53.3	2360	6	ABP56972	Abp56972 E. maxima
548	33	54.1	500	5	ADC10917	Adc10917 Human ext	621	32.5	53.3	2661	6	ABP56958	Abp56958 Bimeria m
549	33	54.1	532	6	ABP77145	Abp77145 N. gonorr	622	32	52.5	12	9	ABE13404	Aeb13404 Tag #5 su
550	33	54.1	532	10	ABE72317	Aee72317 Human car	623	32	52.5	13	5	AAW48823	Aam48823 Endostat1
551	33	54.1	533	4	AY511175	Aay511175 Human bra	624	32	52.5	13	5	ABJ35855	Aab35855 T7 phage
552	33	54.1	533	4	AAU29021	Aau29021 Human PAR	625	32	52.5	14	9	ABJ35855	Abj35855 T7 phage
553	33	54.1	533	10	ABE86985	Aee86985 Human che	626	32	52.5	14	9	ADV58994	Adv58994 G protein
554	33	54.1	535	6	ABU00691	Abu00691 S. pneumo	627	32	52.5	15	9	ADV58450	Adv58450 G protein
555	33	54.1	536	8	ADK47489	Adk47489 Streptoco	628	32	52.5	15	9	ADV59027	Adv59027 G protein
556	33	54.1	540	3	AA511176	Aay511176 Human ute	629	32	52.5	17	3	AAI49353	Aay49353 P. aerugi
557	33	54.1	543	2	AAW70466	Aaw70466 South Afr	630	32	52.5	17	3	AAI49354	Aay49354 P. aerugi
558	33	54.1	548	4	AAU87401	Aau87401 Novel Cen	631	32	52.5	17	8	ADG73765	Adg73765 P. aerugi
559	33	54.1	548	8	ADI54716	Adi54716 Novel hum	632	32	52.5	17	8	ADG73764	Adg73764 P. aerugi
560	33	54.1	550	8	ADS29184	Ads29184 Bacterial	633	32	52.5	18	2	AAI42663	Aay42663 HIV-7 pep
561	33	54.1	560	2	AAW70473	Aaw70473 Girwood	634	32	52.5	44	8	ABO54680	Abos4680 Human gen
562	33	54.1	561	8	ADL14159	Adl14159 Human NF-	635	32	52.5	53	4	ABBI6824	Abbi6824 Human ner
563	33	54.1	561	8	ADP23478	Adp23478 PRO polyp	636	32	52.5	56	7	ADC27504	Adc27504 Western e
564	33	54.1	561	9	ADV73219	Adv73219 Human col	637	32	52.5	59	2	AAW04957	Aaw04957 Partial P
565	33	54.1	565	4	ABG23778	Abg23778 Novel hum	638	32	52.5	59	9	AED55900	Aed55900 Fusarium
566	33	54.1	568	7	ADD22422	Add22422 HLA-B46 T	639	32	52.5	80	3	ABJ58733	Abj58733 Breast an
567	33	54.1	568	7	ADD45724	Add45724 Human Pro	640	32	52.5	85	6	ABP80108	Abp80108 N. gonorr
568	33	54.1	568	7	ADE54650	Ade54650 Human Pro	641	32	52.5	92	4	AAW51744	Aam51744 Human FSH
569	33	54.1	568	7	ADE54646	Ade54646 Human Pro	642	32	52.5	93	7	ABO83220	Abos83220 Pseudomon
570	33	54.1	568	7	ADI15912	Adi15912 Human PP	643	32	52.5	99	2	AAW95288	Aaw95288 Chlamydia
571	33	54.1	568	8	ADO08084	Ado08084 Human pol	644	32	52.5	100	2	AAW95284	Aaw95284 Chlamydia
572	33	54.1	568	8	AEC12341	Aec12341 Human sur	645	32	52.5	105	8	ADN25141	Add25141 Fertility
573	33	54.1	568	9	AEC12726	Aec12726 Human sur	646	32	52.5	105	8	ADN61156	Adn61156 Radish nu
574	33	54.1	582	5	ABP73516	Abp73516 Candida a	647	32	52.5	117	2	AAI37327	Aay37327 Amino aci
575	33	54.1	583	8	ADN23432	Adn23432 Bacterial	648	32	52.5	117	4	ADM20025	Adm20025 Protein e
576	33	54.1	610	5	ABG61879	Abg61879 Prostate	649	32	52.5	121	4	AAU54626	Aau54626 Propionib
577	33	54.1	610	7	ADB75244	Adb75244 Prostate	650	32	52.5	121	6	ABM51145	Abm51145 Propionib
578	33	54.1	610	8	ADL13218	Adl13218 Human ste	651	32	52.5	124	4	ADM19765	Adm19765 Protein e
579	33	54.1	611	8	ADT66701	Adt66701 Murine ca	652	32	52.5	125	3	ABJ40264	Abj40264 Human ORF
580	33	54.1	611	10	AEF79055	Aee79055 Spatzle 3	653	32	52.5	125	5	ABP34931	Abp34931 Human ORF
581	33	54.1	613	5	AAE22074	Aee22074 Gasterost	654	32	52.5	134	2	AAW95286	Aaw95286 Chlamydia
582	33	54.1	614	6	ABM72873	Abm72873 Staphyloc	655	32	52.5	145	2	AAI15094	Aar15094 hcG/hTSH
583	33	54.1	617	8	ADR66942	Adr66942 Human pro	656	32	52.5	149	7	ABO83524	Abos83524 Pseudomon
584	33	54.1	617	8	ADR66044	Adr66044 Human pro	657	32	52.5	154	6	ABU08012	Abu08012 Soybean m
585	33	54.1	637	4	ABG23529	Abg23529 Novel hum	658	32	52.5	158	6	ABU08009	Abu08009 Soybean m
586	33	54.1	641	9	ADY70667	Ady70667 Human nic	659	32	52.5	179	8	ADS28182	Ads28182 Bacterial
587	33	54.1	644	8	ADY23623	Ady23623 Bacterial	660	32	52.5	189	3	AG24503	Aag24503 Arabidops
588	33	54.1	728	7	ADB70233	Adb70233 C. neofor	661	32	52.5	206	8	ADU66592	Adu66592 Human kin
589	33	54.1	803	8	ADJ49938	Adj49938 Oil-a8soc	662	32	52.5	210	7	ADF05490	Adf05490 Bacterial
590	33	54.1	831	8	ADS21941	Ads21941 Bacterial	663	32	52.5	221	7	ABO71523	Abob71523 Pseudomon
591	33	54.1	869	4	ABJ57877	Abj57877 Drosophil	664	32	52.5	226	4	ABG24185	Abg24185 Novel hum
592	33	54.1	872	8	ADJ49939	Adj49939 Oil-a8soc	665	32	52.5	234	8	ADY06333	Ady06333 Plant ful
593	33	54.1	910	5	AAE22072	Aee22072 Gasterost	666	32	52.5	240	2	AAW52309	Aaw52309 Modified
594	33	54.1	951	5	ABP61509	Abp61509 Human NF-	667	32	52.5	244	1	AAI70561	Aap70561 Product o
595	33	54.1	985	4	ABM60399	Abm60399 Drosophil	668	32	52.5	257	9	ADV09395	Adv09395 TrfLCSV ca
596	33	54.1	1037	9	ABE91540	Aeb91540 Microbial	669	32	52.5	261	1	AAP70406	Aap70406 ORF 3 gen
597	33	54.1	1050	5	ABM6245	Abm6245 Bacillus	670	32	52.5	263	8	ADN22777	Adn22777 Bacterial
598	33	54.1	1142	4	ABM62802	Abm62802 Drosophil	671	32	52.5	274	4	AAM93889	Aam93889 Human pol
599	33	54.1	1142	8	ADO08088	Ado08088 Fly polyp	672	32	52.5	274	8	ADL31986	Adl31986 Human pro
600	33	54.1	1191	5	AAU98061	Aau98061 Bacillus	673	32	52.5	277	3	AAI82393	Aay82393 C. tracho
601	33	54.1	1347	8	ADN72391	Adn72391 Thale cre	674	32	52.5	290	4	AAU30224	Aau30224 Novel hum
602	33	54.1	1401	8	ADL13306	Adl13306 Human ste	675	32	52.5	290	7	ABM86508	Abm86508 Rice abio
603	33	54.1	1401	9	ADX07508	Adx07508 Cyclin-de	676	32	52.5	311	5	AAU85734	Aau85734 Human G-P
604	33	54.1	1887	6	ABU20638	Abu20638 Protein e	677	32	52.5	311	5	AAU85735	Aau85735 Human G-P
605	33	54.1	3139	10	ABE39884	Aee39884 Amino aci	678	32	52.5	312	4	AAI71972	Aag71972 Human oif
606	33	54.1	3141	7	ADJ70444	Adj70444 Human hea	679	32	52.5	312	4	AAU24711	Aau24711 Human oif
607	33	54.1	3144	2	AAE58777	Aar58777 Protein e	680	32	52.5	312	5	ABP95827	Abp95827 Human GPC

681	32	52.5	312	5	AAU95659	Human olf	Aau95659	Human olf	754	32	52.5	538	5	AAE23297	Human nec
682	32	52.5	312	5	AAU95776	Human olf	Aau95776	Human olf	755	32	52.5	538	6	ABJ20238	Human IG
683	32	52.5	312	5	AAU85331	G-coupled	Aau85331	G-coupled	756	32	52.5	538	7	ADF13699	Tumor-Aas
684	32	52.5	312	7	ADC85853	Human GPC	Adc85853	Human GPC	757	32	52.5	538	8	ADE86687	Human pol
685	32	52.5	312	7	ADC85621	Human GPC	Adc85621	Human GPC	758	32	52.5	538	9	AED21667	Human can
686	32	52.5	312	8	ADG83522	Human Olf	Adg83522	Human Olf	759	32	52.5	540	8	ADN26066	Bacterial
687	32	52.5	316	6	ADA15561	A. thalia	Ada15561	A. thalia	760	32	52.5	546	8	ABM83047	Human dia
688	32	52.5	316	7	ADC46669	Thalecres	Adc46669	Thalecres	761	32	52.5	555	4	ABG16842	Novel hum
689	32	52.5	316	7	ADDS5808	Thalecres	Adsd5808	Thalecres	762	32	52.5	560	6	ABU09237	Human G-p
690	32	52.5	316	7	ADD30522	Plant yie	Add30522	Plant yie	763	32	52.5	560	6	ADO28902	Human nov
691	32	52.5	316	8	ADI43955	Plant tra	Adi43955	Plant tra	764	32	52.5	562	2	AA41268	Protein h
692	32	52.5	318	6	ABR01613	Human G p	Abro1613	Human G p	765	32	52.5	571	8	ADR09433	Human pro
693	32	52.5	319	9	ABM94945	M. xanthu	Abm94945	M. xanthu	766	32	52.5	572	5	ABP29715	Streptoco
694	32	52.5	323	8	ADSA1881	Bacterial	Adsa1881	Bacterial	767	32	52.5	572	5	ABP26316	Streptoco
695	32	52.5	325	8	ADN18591	Bacterial	Adn18591	Bacterial	768	32	52.5	577	7	ADC06791	Human pro
696	32	52.5	333	4	ABBG3558	Drosophil	Abbg3558	Drosophil	769	32	52.5	577	7	ADC06792	Prostate
697	32	52.5	333	4	ABG16843	Novel hum	Abg16843	Novel hum	770	32	52.5	581	3	AA484322	A human c
698	32	52.5	335	8	ADN23606	Bacterial	Adn23606	Bacterial	771	32	52.5	581	4	ABM65701	Novel pro
699	32	52.5	340	8	ADS24784	Bacterial	Ads24784	Bacterial	772	32	52.5	581	8	ADI29310	Human WAR
700	32	52.5	341	4	AAB96499	Putative	Aab96499	Putative	773	32	52.5	581	8	ADQ15154	Human can
701	32	52.5	341	6	ABU42121	Protein e	Abu42121	Protein e	774	32	52.5	582	4	AAU03896	G protein
702	32	52.5	341	6	ABU39552	Protein e	Abu39552	Protein e	775	32	52.5	584	9	AEB40592	L. pneumo
703	32	52.5	345	2	AAW41166	Metal-reg	Aaw41166	Metal-reg	776	32	52.5	591	5	ABM97017	M. xanthu
704	32	52.5	345	8	ADT06032	Soil rest	Adt06032	Soil rest	777	32	52.5	595	5	ABP65875	Bifidobac
705	32	52.5	346	4	ABB71526	Drosophil	Abb71526	Drosophil	778	32	52.5	599	9	AEB37268	L. pneumo
706	32	52.5	349	3	AA482392	C. tracho	Aay82392	C. tracho	779	32	52.5	605	7	ADF16075	Human alb
707	32	52.5	353	8	ADN21396	Bacterial	Adn21396	Bacterial	780	32	52.5	606	2	AAW84086	Human mem
708	32	52.5	371	6	ABU32109	Protein e	Abu32109	Protein e	781	32	52.5	606	3	AA477489	Human WD-
709	32	52.5	373	2	AAW76362	C. tracho	Aaw76362	C. tracho	782	32	52.5	606	7	ADM05157	Human pro
710	32	52.5	375	8	ADN25794	Bacterial	Adn25794	Bacterial	783	32	52.5	606	8	ADN03646	Antipsori
711	32	52.5	387	7	ABO66852	Klebsiell	AbO66852	Klebsiell	784	32	52.5	606	8	ABM80560	Tumour-as
712	32	52.5	388	8	ADS44090	Bacterial	AdS44090	Bacterial	785	32	52.5	606	8	ADT04108	Novel hum
713	32	52.5	393	2	AAY56759	C. tracho	Aay56759	C. tracho	786	32	52.5	608	9	AEC88087	Human GDN
714	32	52.5	394	1	AA460004	Sequence	Aap60004	Sequence	787	32	52.5	608	8	ADX68346	Plant ful
715	32	52.5	394	2	AAW15149	Chlamydia	Aaw15149	Chlamydia	788	32	52.5	627	4	ABG16844	Novel hum
716	32	52.5	394	2	AAW73141	Chlamydia	Aaw73141	Chlamydia	789	32	52.5	635	7	ABO60660	Klebsiell
717	32	52.5	394	2	AAW57775	Chlamydia	Aaw57775	Chlamydia	790	32	52.5	637	7	ADM04513	Human pro
718	32	52.5	394	2	AA456761	C. tracho	Aay56761	C. tracho	791	32	52.5	637	9	AEC87443	Human cDN
719	32	52.5	394	2	AA457494	Chlamydia	Aay37494	Chlamydia	792	32	52.5	641	6	ABM68514	Photorhab
720	32	52.5	394	3	AA481268	Chlamydia	Aay81268	Chlamydia	793	32	52.5	645	8	ADX68345	Plant ful
721	32	52.5	395	5	ABBS4242	Lactococc	Abb54242	Lactococc	794	32	52.5	647	9	AEB87861	Lipomyces
722	32	52.5	400	5	ABP29950	Streptoco	Abp29950	Streptoco	795	32	52.5	666	8	ADY07866	Plant ful
723	32	52.5	400	8	ADR83914	S. pyogen	Adr83914	S. pyogen	796	32	52.5	672	5	ABP73226	Candida a
724	32	52.5	415	7	ABO83146	Pseudomon	AbO83146	Pseudomon	797	32	52.5	694	7	ADB64315	Human pro
725	32	52.5	422	8	ADS22293	Bacterial	AdS22293	Bacterial	798	32	52.5	694	7	ADC06790	Human pro
726	32	52.5	423	2	AAW27635	Streptoco	Aaw27635	Streptoco	799	32	52.5	694	8	ADQ66699	Novel hum
727	32	52.5	436	4	AAW15541	Peptide #	Aam15541	Peptide #	800	32	52.5	730	7	ADF71289	Novel hum
728	32	52.5	436	4	ABB29374	Peptide #	Abb29374	Peptide #	801	32	52.5	741	6	ADA34611	Acinetoba
729	32	52.5	436	4	AAW55334	Human bra	Aam55334	Human bra	802	32	52.5	743	8	ADV88179	Streptoco
730	32	52.5	436	4	AAW03290	Peptide #	Aam03290	Peptide #	803	32	52.5	743	8	ADV79432	Streptoco
731	32	52.5	436	5	ABG37291	Human pep	Abg37291	Human pep	804	32	52.5	750	5	ABP27691	Streptoco
732	32	52.5	443	4	AAAG82383	S. epider	Aag82383	S. epider	805	32	52.5	758	3	ABM11540	SEN virus
733	32	52.5	444	5	ABP39501	Staphyloc	Abp39501	Staphyloc	806	32	52.5	758	3	ADF70427	Orphan re
734	32	52.5	444	8	ADSO5410	Staphyloc	Adso5410	Staphyloc	807	32	52.5	798	7	ADF70427	Orphan re
735	32	52.5	448	4	ABBG68176	Drosophil	Abb68176	Drosophil	808	32	52.5	818	9	ADX05054	Cat IL4 r
736	32	52.5	454	3	AA482390	C. tracho	Aay82390	C. tracho	809	32	52.5	819	4	ABBG6041	Drosophil
737	32	52.5	457	3	AA482388	C. tracho	Aay82388	C. tracho	810	32	52.5	823	9	ADX05055	Dog IL4 r
738	32	52.5	466	4	AAW25635	Human pro	Aam25635	Human pro	811	32	52.5	842	4	ABB65111	Drosophil
739	32	52.5	466	10	AE888032	Human pro	Aee88032	Human pro	812	32	52.5	882	5	ABP25774	Streptoco
740	32	52.5	471	9	AAW92751	M. xanthu	Abm92751	M. xanthu	813	32	52.5	882	5	ABU46739	Protein e
741	32	52.5	474	4	AAAG62619	Petunia z	Aag62619	Petunia z	814	32	52.5	883	6	AAU06828	S. pneumo
742	32	52.5	474	4	AAAG64525	P. hybrid	Aag64525	P. hybrid	815	32	52.5	883	2	AA408339	S. pneumo
743	32	52.5	474	7	ADG02817	Petunia h	Adg02817	Petunia h	816	32	52.5	883	2	AA490514	Streptoco
744	32	52.5	474	8	ADF38738	Petunia x	Adf38738	Petunia x	817	32	52.5	883	4	AAU38091	Streptoco
745	32	52.5	476	4	ABG20111	Novel hum	Abg20111	Novel hum	818	32	52.5	883	4	AAU37789	Streptoco
746	32	52.5	482	6	ABMG6852	Photorhab	Abm68552	Photorhab	819	32	52.5	883	5	AAU97883	Streptoco
747	32	52.5	484	3	AA482389	C. tracho	Aay82389	C. tracho	820	32	52.5	883	6	ABU00937	S. pneumo
748	32	52.5	494	3	AAAB42297	Human ORF	Ab42297	Human ORF	821	32	52.5	883	6	ABU45915	Protein e
749	32	52.5	496	7	ADG42187	Human bra	Adg42187	Human bra	822	32	52.5	883	8	ADH96823	S. pneumo
750	32	52.5	506	4	AAW25593	Human pro	Aam25593	Human pro	823	32	52.5	883	8	ADH96825	S. pneumo
751	32	52.5	514	3	AA482391	C. tracho	Aay82391	C. tracho	824	32	52.5	883	8	ADK46385	Streptoco
752	32	52.5	514	8	ADN19937	Bacterial	Adn19937	Bacterial	825	32	52.5	883	9	AEC13178	Streptoco
753	32	52.5	534	8	ABM80561	Tumour-as	Abm80561	Tumour-as	826	32	52.5	883	9	AEC13180	Streptoco



827	32	52.5	884	8	ADU69553	Adu69553 S agalact	900	31.5	51.6	2844	7	ADL35963	Adl35963 Human NOV
828	32	52.5	884	8	ADV88336	Adv88336 Streptoco	901	31.5	51.6	2845	5	ABG94631	Abg94631 Human NOV
829	32	52.5	884	8	ADV81756	Adv81756 Streptoco	902	31.5	51.6	2845	7	ADL35967	Adl35967 Human NOV
830	32	52.5	884	8	ADV79589	Adv79589 Streptoco	903	31.5	51.6	2877	5	ABG94630	Abg94630 Human NOV
831	32	52.5	888	8	ADR94329	Adr94329 Novel S.	904	31.5	51.6	2877	7	ADL35965	Adl35965 Human NOV
832	32	52.5	888	7	Aea58199	Aea58199 Streptoco	905	31.5	51.6	2995	7	ADL35961	Adl35961 Human NOV
833	32	52.5	912	7	ADD14183	Add14183 Human src	906	31.5	51.6	3415	9	AEE04798	Aee04798 Cancer-as
834	32	52.5	926	5	ABP69492	Abp69492 Human pol	907	31	50.8	9	3	AAV69597	AAV69597 Monoclonal
835	32	52.5	926	8	ADR40159	Adr40159 Human pro	908	31	50.8	9	3	AAV50888	AAV50888 Antibody
836	32	52.5	926	8	ADP54550	Adp54550 Human PRO	909	31	50.8	9	9	ADU70618	Adu70618 Human hep
837	32	52.5	926	9	ADX05516	Adx05516 Cyclin-de	910	31	50.8	10	9	ADV59607	Adv59607 G protein
838	32	52.5	926	9	ADY15498	Ady15498 PRO polyp	911	31	50.8	12	2	Aaw38089	Aaw38089 PPPY mot
839	32	52.5	931	6	ABU31995	Abu31995 Protein e	912	31	50.8	12	7	ADB49350	Adb49350 Biocinyla
840	32	52.5	951	4	AAU34872	Aau34872 E. coli c	913	31	50.8	14	9	ADV59626	Adv59626 G protein
841	32	52.5	951	4	AAU38260	Aau38260 Salmonell	914	31	50.8	15	9	ADU71085	Adu71085 Human hep
842	32	52.5	951	6	ABU15050	Abu15050 Protein e	915	31	50.8	16	5	ABV79701	Abv79701 Integrin
843	32	52.5	951	6	ABU48273	Abu48273 Protein e	916	31	50.8	16	9	ADV59623	Adv59623 G protein
844	32	52.5	951	6	ABU28309	Abu28309 Protein e	917	31	50.8	16	9	ADZ68070	Adz68070 MAP kinase
845	32	52.5	953	6	ABU49570	Abu49570 Protein e	918	31	50.8	17	3	AAB25992	Aab25992 Human Igg
846	32	52.5	962	6	ABU41155	Abu41155 Protein e	919	31	50.8	17	4	AAU16717	Aau16717 Peptide E
847	32	52.5	965	7	ABU49983	Abu49983 Protein e	920	31	50.8	17	5	ABJ00247	Abj00247 Human Igg
848	32	52.5	967	7	ABO62249	Ab062249 Klebsiell	921	31	50.8	23	7	ADB87673	Adb87673 Human neu
849	32	52.5	975	7	ADF05470	Adf05470 Bacterial	922	31	50.8	27	2	AAW14660	Aaw14660 Integrin
850	32	52.5	975	6	ABM68329	Abm68329 Photorhab	923	31	50.8	27	5	AAM50485	Aam50485 Integrin
851	32	52.5	982	8	ADS29846	Ads29846 Bacterial	924	31	50.8	27	5	AAM50503	Aam50503 Integrin
852	32	52.5	989	7	ADM29380	Adm29380 Human nov	925	31	50.8	27	5	AAM50504	Aam50504 Integrin
853	32	52.5	1006	6	ADP80935	Adp80935 N. gonorr	926	31	50.8	37	7	ADC88007	Adc88007 Ribosomal
854	32	52.5	1010	9	ADB91379	Adb91379 Microbial	927	31	50.8	38	5	ABP28976	Abp28976 Streptoco
855	32	52.5	1019	8	ADN22389	Adn22389 Bacterial	928	31	50.8	40	4	AAB66762	Aab66762 Beta2 sub
856	32	52.5	1047	5	ABG61533	Abg61533 Human tra	929	31	50.8	40	5	ABG60572	Abg60572 Selective
857	32	52.5	1047	7	ADE63021	Ad663021 Human pro	930	31	50.8	40	5	ABV79719	Abv79719 Integrin
858	32	52.5	1047	7	ADE63017	Ade63017 Human Pro	931	31	50.8	42	3	AAV90537	AAV90537 Conus sul
859	32	52.5	1047	8	ADQ88701	Adq88701 Human ATP	932	31	50.8	42	3	AAV90534	AAV90534 Conus atr
860	32	52.5	1082	9	AEE02346	Aee02346 Human her	933	31	50.8	42	3	AAV32353	AAV32353 Human Clq
861	32	52.5	1093	2	AAR41001	Aar41001 Human myo	934	31	50.8	46	2	AAR80964	Aar80964 Integrin
862	32	52.5	1132	4	ABG07312	Abg07312 Novel hum	935	31	50.8	46	2	AAW65886	AAW65886 Cytoplasm
863	32	52.5	1272	6	ABP72190	Abp72190 Plasmodiu	936	31	50.8	46	2	AAW48690	AAW48690 Amino aci
864	32	52.5	1429	5	ABP35654	Abp35654 Fungal ZB	937	31	50.8	46	2	AAV43097	AAV43097 Human int
865	32	52.5	1433	5	ABP35624	Abp35624 Fungal ZB	938	31	50.8	46	3	AAB26128	Aab26128 Beta2-int
866	32	52.5	1610	7	ADB46133	Adb46133 Plasmodiu	939	31	50.8	46	8	ABO54611	Ab054611 Human gen
867	32	52.5	1804	6	ABR53154	AbR53154 Protein s	940	31	50.8	47	3	AAU62166	Aau62166 Propionib
868	32	52.5	1804	7	ADK63138	Adk63138 Disease t	941	31	50.8	54	4	AAU62166	Aau62166 Propionib
869	32	52.5	2000	4	ABG61183	Abg61183 Myxoma vi	942	31	50.8	54	6	ABM58685	Abm58685 Propionib
870	32	52.5	2197	8	ADK16573	Adk16573 Nanoarcha	943	31	50.8	57	5	ABB81140	Abb81140 Human ret
871	32	52.5	4655	2	AAW43312	Aaw43312 Human pla	944	31	50.8	63	6	AAU42527	Aau42527 Propionib
872	31.5	51.6	85	2	AAW55353	Aaw55353 H. pylori	945	31	50.8	63	6	ABM39046	Abm39046 Propionib
873	31.5	51.6	244	4	AAB98655	Aab98655 Caspase-7	946	31	50.8	71	5	AAV79388	AAV79388 EGF-like
874	31.5	51.6	303	2	AAW15262	Aaw15262 Apoptotic	947	31	50.8	71	5	AAE17041	Aae17041 Human G p
875	31.5	51.6	303	2	AAW15247	Aaw15247 Cysteine	948	31	50.8	71	8	ADL14217	Adl14217 Novel hum
876	31.5	51.6	303	2	AAV21721	Aay21721 Amino aci	949	31	50.8	77	2	AAR88749	Aar88749 S. aureus
877	31.5	51.6	303	4	AAE00604	Aae00604 Human cas	950	31	50.8	77	4	AAU63452	Aau63452 Propionib
878	31.5	51.6	303	5	ABJ01222	Abj01222 Human cas	951	31	50.8	77	5	ABP02587	Abp02587 Human ORF
879	31.5	51.6	303	5	ABO9299	Ab09299 Human cas	952	31	50.8	77	6	ABM59971	Abm59971 Propionib
880	31.5	51.6	303	5	ABO9297	Ab09297 Human cas	953	31	50.8	82	7	ADF30910	Adf30910 Soil meta
881	31.5	51.6	303	6	AAO19868	Aao19868 Bacteriop	954	31	50.8	84	3	AAG61260	Aag61260 Arabidops
882	31.5	51.6	303	8	ADO16852	Ado16852 Human cas	955	31	50.8	84	3	AAG57207	Aag57207 Arabidops
883	31.5	51.6	305	9	ADV90184	Adv90184 Human pro	956	31	50.8	85	2	AAW04170	Aaw04170 Flea calr
884	31.5	51.6	305	9	ADV69228	Adv69228 Human Cys	957	31	50.8	88	4	ABG25271	Abg25271 Novel hum
885	31.5	51.6	327	10	Aef11566	Aef11566 Tomato me	958	31	50.8	89	5	ADK35323	Adk35323 Novel hum
886	31.5	51.6	330	4	ABB69802	Abb69802 Drosophil	959	31	50.8	91	3	AAG57206	Aag57206 Arabidops
887	31.5	51.6	330	8	ADS69506	Ads69506 Drosophil	960	31	50.8	92	4	AAM51745	Aam51745 Human FSH
888	31.5	51.6	336	5	ABO9300	Ab09300 Human cas	961	31	50.8	96	3	AAG57205	Aag57205 Arabidops
889	31.5	51.6	341	2	AAR95830	Aar95830 Human int	962	31	50.8	96	3	AAG61259	Aag61259 Arabidops
890	31.5	51.6	1118	4	AAM50209	Aam50209 Human fib	963	31	50.8	99	8	ADR96565	Adr96565 Novel S.
891	31.5	51.6	1600	5	ABG66678	Abg66678 Human nov	964	31	50.8	99	9	AEA60435	Aea60435 Streptoco
892	31.5	51.6	1716	9	AEE04794	Aee04794 Cancer-as	965	31	50.8	101	2	AAR90939	Aar90939 D3D4 poly
893	31.5	51.6	1910	7	AEE04796	Aee04796 Cancer-as	966	31	50.8	102	3	AAG57676	Aag57676 Arabidops
894	31.5	51.6	2676	9	ADL35971	Adl35971 Human NOV	967	31	50.8	102	3	AAG10138	Aag10138 Arabidops
895	31.5	51.6	2695	5	ABG94632	Abg94632 Human NOV	968	31	50.8	103	2	AAV35877	Aav35877 C. pneumo
896	31.5	51.6	2695	7	ADL35969	Adl35969 Human NOV	969	31	50.8	109	4	ABG10963	Abg10963 Novel hum
897	31.5	51.6	2757	5	ABG94633	Abg94633 Human NOV	970	31	50.8	113	2	AAR90937	Aar90937 D3D4 poly
898	31.5	51.6	2809	5	AAG66169	Aag66169 Human fib	971	31	50.8	113	4	ABB62220	Abb62220 Drosophil
899	31.5	51.6	2844	5	ABG94629	Abg94629 Human NOV	972	31	50.8	115	5	ADK36270	Adk36270 Novel hum

973 31 50.8 116 5 ABP06792  
974 31 50.8 117 5 ADK34293  
975 31 50.8 123 8 ADX95695  
976 31 50.8 124 2 AAR90940  
977 31 50.8 126 8 ADS24055  
978 31 50.8 127 8 ADY12815  
979 31 50.8 128 2 AAR90941  
980 31 50.8 129 3 AAG10137  
981 31 50.8 130 2 AAR58806  
982 31 50.8 130 2 AAR58807  
983 31 50.8 130 2 AAR91202  
984 31 50.8 130 2 AAR91201  
985 31 50.8 131 2 AAR95447  
986 31 50.8 131 8 ADX95773  
987 31 50.8 132 4 AAU29691  
988 31 50.8 133 4 AAU50531  
989 31 50.8 133 6 ABM47050  
990 31 50.8 134 2 AAU29089  
991 31 50.8 134 2 AAY29088  
992 31 50.8 135 3 AAG57675  
993 31 50.8 137 2 AAY44012  
994 31 50.8 140 3 AAG57674  
995 31 50.8 141 4 AAU08682  
996 31 50.8 141 7 ADB32031  
997 31 50.8 145 4 AAU30715  
998 31 50.8 147 8 ADO26501  
999 31 50.8 153 3 AAG10136  
1000 31 50.8 154 8 ADY11312

ALIGNMENTS

RESULT 1  
ADR88212  
ID ADR88212 standard; peptide; 10 AA.  
XX AC ADR88212;  
XX DT 18-NOV-2004 (first entry)  
XX DE Human heparanase epitope pep38.  
XX KW Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;  
KW autoimmune disorder; cancer; angiogenesis; metastatic disease;  
KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;  
KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;  
KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;  
KW human; heparanase; enzyme; epitope.  
XX OS Homo sapiens.  
XX PN US2004170631-A1.  
XX PD 02-SEP-2004.  
XX PF 28-NOV-2003; 2003US-00722502.  
XX PR 02-SEP-1997; 97US-00922170.  
PR 01-MAY-1998; 98US-00071739.  
PR 04-NOV-1998; 98US-00186200.  
PR 19-FEB-2003; 2003US-00368044.  
PR 22-AUG-2003; 2003US-00645659.  
XX (YACO/) YACOBY-ZEEVI O.  
PA (PERE/) PERETZ T.  
PA (MIRO/) MIRON D.  
PA (SHLO/) SHLOMI Y.  
PA (PECK/) PECKER I.  
PA (AYAL/) AYAL-HERSHKOVITZ M.  
PA (FEIN/) FEINSTEIN E.  
PA (VGEL/) VAN GELDER J M.  
PA (VLOD/) VLODAVSKY I.

PA (FRIE/) FRIEDMANN Y.  
XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;  
XX Friedmann Y;  
XX WPI; 2004-625084/60.  
XX Targeted drug delivery to a heparanase-expressing tissue of a patient,  
PT useful for treating heparanase-associated conditions such as inflammation  
PT or cancer, comprises administering a drug and an anti-heparanase antibody  
XX complex.  
XX Claim 7; SEQ ID NO 6; 58pp; English.  
XX The invention relates to a method of targeted drug delivery to a tissue  
CC of a patient, the tissue expressing heparanase. The method comprises  
CC providing a complex of a drug directly or indirectly linked to an anti-  
CC heparanase antibody, and administering the complex to the patient. In the  
CC targeted drug delivery, the antibody comprises an antibody or its portion  
CC capable of specifically binding to at least one epitope of a heparanase  
CC protein. The composition and methods of the invention are useful for  
CC diagnosing, preventing or treating conditions associated with heparanase  
CC catalytic activity (e.g. an inflammatory disorder, wound, scar,  
CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell  
CC proliferation, invasion of circulating tumour cells and metastatic  
CC disease), for purifying heparanase, or for developing drugs for those  
CC heparanase-associated conditions. The vasculopathy is atherosclerosis,  
CC restenosis or aneurysm. The cancerous condition is a solid cancer or a  
CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy  
CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous  
CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous  
CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,  
CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and  
CC multiple myeloma. The solid cancer is selected from tumours in lip and  
CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,  
CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,  
CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of  
CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue  
CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,  
CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic  
CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary  
CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,  
CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,  
CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,  
CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's  
CC sarcoma. The present sequence is human heparanase epitope.  
XX SQ Sequence 10 AA;

Query Match 100.0%; Score 61; DB 8; Length 10;  
Best Local Similarity 100.0%; Pred. NO. 0.0016;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTNTDNPYK 10  
| | | | | | | | | |  
Db 1 CTNTDNPYK 10

RESULT 2  
ADT78179  
ID ADT78179 standard; peptide; 10 AA.  
XX AC ADT78179;  
XX DT 13-JAN-2005 (first entry)  
XX DE Functional peptide epitope of human heparanase, pep38.

XX Antibody; epitope; heparanase; pathological condition; angiogenesis;  
KW cell proliferation; cancerous condition; tumour cell invasion;  
KW metastatic disease; heparanase-related disorder; inflammatory disorder;  
KW wound; scar; vasculopathy; autoimmune condition; renal disease;

KW cytostatic; antiinflammatory; vulnery; antiarteriosclerotic;  
 KW vasotrophic; immunosuppressive; nephrotropic; antidiabetic; human.  
 XX

OS Homo sapiens.

XX US2004213789-A1.

XX PN 28-OCT-2004.

XX PF 22-AUG-2003; 2003US-00645659.

XX PR 02-SEP-1997; 97US-00922170.

XX PR 01-MAY-1998; 98US-00071739.

XX PR 04-NOV-1998; 98US-00186200.

XX PR 19-FEB-2003; 2003US-00368044.

XX (YACO/) YACOBY-ZEEVI O.

PA (PERE/) PERETZ T.

PA (MIRO/) MIRON D.

PA (SHLO/) SHLOMI Y.

PA (PECK/) PECKER I.

PA (AYAL/) AYAL-HERSHKOVITZ M.

PA (FEIN/) FEINSTEIN E.

PA (GELD/) GELDER J M V.

PA (VLOD/) VLODAVSKY I.

PA (FRIE/) FRIEDMANN Y.

XX

PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;

PI Ayal-Hershkovitz M, Feinstein E, Gelder JMV, Vlodavsky I;

PI Friedmann Y;

XX WPI; 2004-774790/76.

XX New neutralizing monoclonal anti-heparanase antibodies, useful for  
 PT detecting, treating or preventing cancer, inflammatory or autoimmune  
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.  
 XX

PS Claim 39; SEQ ID NO 6; 68pp; English.

XX The invention relates to an isolated antibody or antibody portion capable  
 CC of specifically binding to or elicited by at least one epitope of a  
 CC heparanase protein, where the heparanase protein is at least 60%  
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and  
 CC where at least one epitope comprises a sequence at least 70% homologous  
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)  
 CC a hybridoma cell line comprising a cell line for producing the monoclonal  
 CC antibody, (b) a method for detecting, treating or preventing a  
 CC pathological condition or a heparanase-related disorder or condition in a  
 CC subject, (c) a method for monitoring the state of a heparanase-related  
 CC disorder or condition in a subject, and (d) a pharmaceutical composition  
 CC comprising the isolated anti-heparanase antibody or antibody portion and  
 CC a pharmaceutical carrier. The antibody, methods, and composition are  
 CC useful for detecting, treating, preventing or monitoring a pathological  
 CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition  
 CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,  
 CC or prostate cancer), minor cell proliferation, invasion of circulating  
 CC tumor cells, or a metastatic disease, or a heparanase-related disorder  
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy  
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or  
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,  
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell  
 CC carcinoma) in a mammal. This sequence represents a functional peptide  
 CC epitope of human heparanase.

XX SQ Sequence 10 AA;

Query Match 100.0%; Score 61; DB 8; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.0016;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10

DB 1 CTNTDNPYK 10

RESULT 3  
 AEA42428

XX ID AEA42428 standard; peptide; 10 AA.

XX AC AEA42428;

XX DT 28-JUL-2005 (first entry)

XX DE Human heparanase epitope peptide SEQ ID NO:6.

XX antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;  
 KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;  
 KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;  
 KW angiogenesis disorder; cancer; tumor; metastasis; epitope.

XX Homo sapiens.

XX PN AU2004201462-A1.

XX PD 06-MAY-2004.

XX PF 08-APR-2004; 2004AU-00201462.

XX PR 08-APR-2004; 2004AU-00201462.

XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX Vlodavsky I, Pecker I, Miron M, Gilboa A, Miron D, Moskowitz H;  
 PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;  
 PI Feinstein E;

XX WPI; 2005-173343/19.

XX Novel isolated antibody capable of specifically binding to epitope of  
 PT heparanase protein, useful for preventing and treating heparanase-related  
 PT disorder such as inflammatory disorder, scars, autoimmune conditions or  
 PT angiogenesis.

XX Claim 7; SEQ ID NO 6; 260pp; English.

XX The invention relates to an isolated antibody or its portion (I) capable  
 CC of specifically binding to an epitope of a heparanase protein. Also  
 CC described: (1) a cell line (II) for producing a monoclonal antibody or  
 CC its portion, comprising a cell line for producing (I); (2) a  
 CC pharmaceutical composition comprising (I) and a carrier; and (3) an  
 CC affinity medium (III) for binding human heparanase polypeptides,  
 CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)  
 CC useful for treating a subject suffering from a pathological condition,  
 CC which involves administering (I) to the subject. (I) is useful for  
 CC preventing and treating heparanase-related disorder or condition chosen  
 CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune  
 CC condition, angiogenesis, cell proliferation, cancerous condition, tumor  
 CC cell proliferation, invasion of circulating tumor cells and metastatic  
 CC disease. (I) is useful for detecting the presence of heparanase  
 CC polypeptide in a sample. (I) is useful for detecting heparanase-related  
 CC disease or condition in a subject such as vertebrate, preferably mammal  
 CC e.g., human. The heparanase-related disorder or condition further  
 CC includes renal disease or disorder chosen from diabetic nephropathy,  
 CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome  
 CC and renal cell carcinoma. The present sequence represents a human  
 CC heparanase epitope peptide, which is used in the exemplification of the  
 CC present invention.

XX SQ Sequence 10 AA;

Query Match 100.0%; Score 61; DB 9; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.0016;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10



PT for inducing an immune response or for treating metastatic tumors.

PS Disclosure; SEQ ID NO 772; 269pp; English.

XX The invention relates to a novel heparanase peptide that binds to a human  
CC leukocyte antigen (HLA) molecule, where the peptide is a nonapeptide, or  
CC its functional derivative. A peptide of the invention has immunostimulant  
CC and cytostatic activity, and is used in a vaccine. The heparinase peptide  
CC is useful for preparing a medicament which induces an immune response or  
CC for treating metastatic tumors. The present sequence represents a  
XX heparinase peptide of the invention.

XX Sequence 15 AA;

Query Match 100.0%; Score 61; DB 9; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.0024;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10  
| | | | | | | | | |  
Db 3 CTNTDNPYK 12

RESULT 7  
ADU71245  
ID ADU71245 standard; peptide; 15 AA.

XX AC ADU71245;

XX DT 10-FEB-2005 (first entry)

XX DE Human heparanase peptide SEQ ID NO:930.

XX KW enzyme; heparinase; vaccine; human leukocyte antigen; HLA;  
KW immunostimulant; cytostatic; immune disorder; metastasis.

XX OS Homo sapiens.

XX PN EP1479764-A1.

XX PD 24-NOV-2004.

XX PF 19-MAY-2003; 2003EP-00011038.

XX PR 19-MAY-2003; 2003EP-00011038.

XX PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.

XX PA (UYHE-) UNIV HEIDELBERG RUPRECHT-KARLS.

XX PI Schirmacher V, Beckhove P, Sommerfeldt N;

XX DR WPI; 2005-014847/02.

XX PT New heparanase nonapeptide that binds to a human leukocyte antigen (HLA)  
PT molecule or its functional derivative, useful for preparing a medicament  
PT for inducing an immune response or for treating metastatic tumors.

XX PS Disclosure; SEQ ID NO 930; 269pp; English.

XX CC The invention relates to a novel heparanase peptide that binds to a human  
CC leukocyte antigen (HLA) molecule, where the peptide is a nonapeptide, or  
CC its functional derivative. A peptide of the invention has immunostimulant  
CC and cytostatic activity, and is used in a vaccine. The heparinase peptide  
CC is useful for preparing a medicament which induces an immune response or  
CC for treating metastatic tumors. The present sequence represents a  
CC heparinase peptide of the invention.

XX Sequence 15 AA;

Query Match 100.0%; Score 61; DB 9; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.0024;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10  
| | | | | | | | | |  
Db 4 CTNTDNPYK 13

RESULT 8  
ADR88207

ID ADR88207 standard; protein; 386 AA.

XX AC ADR88207;

XX DT 18-NOV-2004 (first entry)

XX DE Human mature heparanase 45 kDa major subunit.

XX KW Targeted drug delivery; inflammatory disorder; wound; scar;  
KW vasculopathy; autoimmune disorder; cancer; angiogenesis;  
KW metastatic disease; atherosclerosis; restenosis; aneurysm; solid cancer;  
KW non-solid cancer; haematopoietic malignancy; lymphocytic leukaemia;  
KW myelogenous leukaemia; Hodgkin's disease; multiple myeloma;  
KW haemangiosarcoma; Kaposi's sarcoma; human; heparanase; enzyme.

XX OS Homo sapiens.

XX PN US2004170631-A1.

XX PD 02-SEP-2004.

XX PF 28-NOV-2003; 2003US-00722502.

XX PR 02-SEP-1997; 97US-00922170.

XX PR 01-MAY-1998; 98US-00071739.

XX PR 04-NOV-1998; 98US-00186200.

XX PR 19-FEB-2003; 2003US-00368044.

XX PR 22-AUG-2003; 2003US-00645659.

XX PA (YACO/) YACOBY-ZEEVI O.

XX PA (PERE/) PERETZ T.

XX PA (MIRO/) MIRON D.

XX PA (SHLO/) SHLOMI Y.

XX PA (PECK/) PECKER I.

XX PA (AYAL/) AYAL-HERSHKOVITZ M.

XX PA (FEIN/) FEINSTEIN E.

XX PA (VGEL/) VAN GELDER J M.

XX PA (VLOD/) VLODAVSKY I.

XX PA (FRIE/) FRIEDMANN Y.

XX PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;

PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;

PI Friedmann Y;

XX WPI; 2004-625084/60.

XX PT Targeted drug delivery to a heparanase-expressing tissue of a patient,  
PT useful for treating heparanase-associated conditions such as inflammation  
PT or cancer, comprises administering a drug and an anti-heparanase antibody  
PT complex.

XX PS Claim 2; SEQ ID NO 1; 59pp; English.

XX CC The invention relates to a method of targeted drug delivery to a tissue  
CC of a patient, the tissue expressing heparanase. The method comprises  
CC providing a complex of a drug directly or indirectly linked to an anti-  
CC heparanase antibody, and administering the complex to the patient. In the  
CC targeted drug delivery, the antibody comprises an antibody or its portion  
CC capable of specifically binding to at least one epitope of a heparanase  
CC protein. The composition and methods of the invention are useful for  
CC diagnosing, preventing or treating conditions associated with heparanase  
CC catalytic activity (e.g. an inflammatory disorder, wound, scar,  
CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell  
CC proliferation, invasion of circulating tumour cells and metastatic  
CC disease), for purifying heparanase, or for developing drugs for those  
CC heparanase-associated conditions. The vasculopathy is atherosclerosis,

CC restenosis or aneurysm. The cancerous condition is a solid cancer or a  
 CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy  
 CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous  
 CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous  
 CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,  
 CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and  
 CC multiple myeloma. The solid cancer is selected from tumours in lip and  
 CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,  
 CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,  
 CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of  
 CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue  
 CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,  
 CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic  
 CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary  
 CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,  
 CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,  
 CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,  
 CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's  
 CC sarcoma. The present sequence is the 45 kDa major subunit of human mature  
 CC heparanase.  
 XX  
 SQ Sequence 386 AA;  
 Query Match 100.0%; Score 61; DB 8; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 0.06;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTNTDNPYK 10  
 DB 280 CTNTDNPYK 289  
 ADT78174;  
 AC  
 XX  
 XX 13-JAN-2005 (first entry)  
 XX  
 DE 45kDa subunit of mature processed human heparanase dimer.  
 KW Antibody; epitope; heparanase; pathological condition; angiogenesis;  
 KW cell proliferation; cancerous condition; tumour cell invasion;  
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;  
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;  
 KW cytostatic; antiinflammatory; vulnery; arteriosclerotic;  
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.  
 XX  
 OS Homo sapiens.  
 XX  
 XX US2004213789-A1.  
 XX  
 XX 28-OCT-2004.  
 XX  
 XX 22-AUG-2003; 2003US-00645659.  
 XX  
 PR 02-SEP-1997; 97US-00922170.  
 PR 01-MAY-1998; 98US-00071739.  
 PR 04-NOV-1998; 98US-00186200.  
 PR 19-FEB-2003; 2003US-00368044.  
 XX  
 XX (YACO/) YACOBY-ZEEVI O.  
 XX (PERE/) PERETZ T.  
 XX (MIRO/) MIRON D.  
 XX (SHLO/) SHLOMI Y.  
 XX (PECK/) PECKER I.  
 XX (AYAL/) AYAL-HERSHKOVITZ M.  
 XX (FEIN/) FEINSTEIN E.  
 XX (GELD/) GELDER J M V.  
 XX (VLOD/) VLODAVSKY I.  
 XX (FRIE/) FRIEDMANN Y.  
 XX

PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
 PI Ayal-Hershkovitz M, Feinstein E, Gelder JMV, Vlodavsky I;  
 XX Friedmann Y;  
 DR WPI; 2004-774790/76.  
 XX  
 PT New neutralizing monoclonal anti-heparanase antibodies, useful for  
 PT detecting, treating or preventing cancer, inflammatory or autoimmune  
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.  
 XX  
 PS Claim 5; SEQ ID NO 1; 68pp; English.  
 XX  
 CC The invention relates to an isolated antibody or antibody portion capable  
 CC of specifically binding to or elicited by at least one epitope of a  
 CC heparanase protein, where the heparanase protein is at least 60%  
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and  
 CC where at least one epitope comprises a sequence at least 70% homologous  
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)  
 CC a hybridoma cell line comprising a cell line for producing the monoclonal  
 CC antibody, (b) a method for detecting, treating or preventing a  
 CC pathological condition or a heparanase-related disorder or condition in a  
 CC subject, (c) a method for monitoring the state of a heparanase-related  
 CC disorder or condition in a subject, and (d) a pharmaceutical composition  
 CC comprising the isolated anti-heparanase antibody or antibody portion and  
 CC a pharmaceutical carrier. The antibody, methods, and composition are  
 CC useful for detecting, treating, preventing or monitoring a pathological  
 CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition  
 CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,  
 CC or prostate cancer), minor cell proliferation, invasion of circulating  
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder  
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy  
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or  
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,  
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell  
 CC carcinoma) in a mammal. This sequence represents the 45kDa subunit of  
 CC mature processed human heparanase dimer.  
 XX  
 SQ Sequence 386 AA;  
 Query Match 100.0%; Score 61; DB 8; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 0.06;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTNTDNPYK 10  
 DB 280 CTNTDNPYK 289  
 ADT78174;  
 AC  
 XX  
 XX 13-JAN-2005 (first entry)  
 XX  
 DE 45kDa subunit of mature processed human heparanase dimer.  
 KW Antibody; epitope; heparanase; pathological condition; angiogenesis;  
 KW cell proliferation; cancerous condition; tumour cell invasion;  
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;  
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;  
 KW cytostatic; antiinflammatory; vulnery; arteriosclerotic;  
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.  
 XX  
 OS Homo sapiens.  
 XX  
 XX US2004213789-A1.  
 XX  
 XX 28-OCT-2004.  
 XX  
 XX 22-AUG-2003; 2003US-00645659.  
 XX  
 PR 02-SEP-1997; 97US-00922170.  
 PR 01-MAY-1998; 98US-00071739.  
 PR 04-NOV-1998; 98US-00186200.  
 PR 19-FEB-2003; 2003US-00368044.  
 XX  
 XX (YACO/) YACOBY-ZEEVI O.  
 XX (PERE/) PERETZ T.  
 XX (MIRO/) MIRON D.  
 XX (SHLO/) SHLOMI Y.  
 XX (PECK/) PECKER I.  
 XX (AYAL/) AYAL-HERSHKOVITZ M.  
 XX (FEIN/) FEINSTEIN E.  
 XX (GELD/) GELDER J M V.  
 XX (VLOD/) VLODAVSKY I.  
 XX (FRIE/) FRIEDMANN Y.  
 XX

```

XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
XX PI Van-Gelder JM, Miron D;
XX DR WPI; 2005-182203/19.
XX PT Regulating heparanase activity, useful for treating heparanase-associated
XX PT diseases (e.g. cancer, inflammation, cardiovascular diseases, heparanase
XX PT neurological diseases or viral diseases) comprises modulating heparanase
XX PT activation.
XX PS Claim 55; SEQ ID NO 33; 211pp; English.
XX CC The invention relates to a method of regulating heparanase activity in a
XX CC tissue or regulating a biological process depending at least in part on
XX CC heparanase activity comprising modulating heparanase activation. The
XX CC invention also relates to methods of treating a heparanase- or heparin
XX CC binding protein-associated disease or disorder in a subject, a
XX CC pharmaceutical composition for use in the treatment of a heparanase-
XX CC associated disease or disorder comprising a therapeutic amount of an
XX CC agent capable of modulating heparanase activation and a pharmaceutical
XX CC carrier or diluent, a method of identifying a protease activator of
XX CC heparanase, a protease substrate mimetic comprising a peptide
XX CC representing a subset or all substrate residues or cleavage sites of
XX CC human heparanase or an equivalent non-human heparanase, a method of
XX CC producing active heparanase and a method of modulating an adhesion
XX CC activity of heparanase. The composition and methods are useful for
XX CC modulating heparanase activation and for treating heparanase-associated
XX CC diseases or disorders such as cancer, inflammation, cardiovascular
XX CC diseases, neurological diseases or viral infections. This sequence
XX CC represents a heparanase inhibitor protein used in the scope of the
XX CC invention.
XX SQ Sequence 386 AA;

Query Match 100.0%; Score 61; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 0.06;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
Db 280 CTNTDNPYK 289

RESULT 11
ADZ18995
ID ADZ18995 standard; protein; 386 AA.
XX AC ADZ18995;
XX DT 16-JUN-2005 (first entry)
XX DE Human heparanase consensus cleavage site #2.
XX KW Enzyme engineering; heparanase; metastasis; autoimmune disease;
XX KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
XX KW immunosuppressive; enzyme.
XX OS Homo sapiens.
XX PN WO2005030962-A1.
XX PD 07-APR-2005.
XX PF 17-SEP-2004; 2004WO-EP010517.
XX PR 26-SEP-2003; 2003US-0506479P.
XX PR 20-JAN-2004; 2004US-0537729P.
XX PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
XX PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;

XX DR WPI; 2005-273382/28.
XX PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a
XX PT nucleotide sequence that encodes mammalian heparanase protein and has two
XX PT consensus cleavage sites located between specific nucleotide encoding
XX PT residues.
XX PS Disclosure; SEQ ID NO 16; 65pp; English.
XX CC The invention relates to a synthetic nucleic acid molecule that encodes
XX CC mammalian heparanase protein, where the nucleic acid comprises two
XX CC consensus cleavage sites recognized by endoproteinase. The sequences are
XX CC useful for expressing mammalian heparanase in non-mammalian cells and in
XX CC inhibitor screening assays for the development of therapeutics or
XX CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
XX CC and/or inflammation. This sequence represents a human heparanase
XX CC consensus cleavage site used in the scope of the invention.
XX SQ Sequence 386 AA;

Query Match 100.0%; Score 61; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 0.06;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10
Db 280 CTNTDNPYK 289

RESULT 12
AEA42423
ID AEA42423 standard; protein; 386 AA.
XX AC AEA42423;
XX DT 28-JUL-2005 (first entry)
XX DE Human mature heparanase dimer 45 kDa subunit SEQ ID NO:1.
XX KW antibody; heparanase; antiinflammatory; vulnary; immunosuppressive;
XX KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;
XX KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;
XX KW angiogenesis disorder; cancer; tumor; metastasis.
XX OS Homo sapiens.
XX PN AU2004201462-A1.
XX PD 06-MAY-2004.
XX PF 08-APR-2004; 2004AU-00201462.
XX PR 08-APR-2004; 2004AU-00201462.
XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.
XX PI Vlodavsky I, Pecker I, Miron M, Gilboa A, Miron D, Moskowitz H;
XX PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;
XX PI Feinstein E;
XX DR WPI; 2005-173343/19.
XX PT Novel isolated antibody capable of specifically binding to epitope of
XX PT heparanase protein, useful for preventing and treating heparanase-related
XX PT disorder such as inflammatory disorder, scars, autoimmune conditions or
XX PT angiogenesis.
XX PS Claim 2; SEQ ID NO 1; 260pp; English.
XX CC The invention relates to an isolated antibody or its portion (I) capable
XX CC of specifically binding to an epitope of a heparanase protein. Also

```



described: (1) a cell line (II) for producing a monoclonal antibody or its portion, comprising a cell line for producing (I); (2) a pharmaceutical composition comprising (I) and a carrier; and (3) an affinity medium (III) for binding human heparanase polypeptides, comprising (I) immobilized to a chemically inert, insoluble carrier. (I) is useful for treating a subject suffering from a pathological condition, which involves administering (I) to the subject. (I) is useful for preventing and treating heparanase-related disorder or condition chosen from inflammatory disorder, wound, scar, vasculopathy, autoimmune condition, angiogenesis, cell proliferation, cancerous condition, tumor cell proliferation, invasion of circulating tumor cells and metastatic disease. (I) is useful for detecting the presence of heparanase polypeptide in a sample. (I) is useful for detecting heparanase-related disease or condition in a subject such as vertebrate, preferably mammal e.g., human. The heparanase-related disorder or condition further includes renal disease or disorder chosen from diabetic nephropathy, glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome and renal cell carcinoma. The present sequence represents the 45 kDa subunit of the human mature processed heparanase dimer, which is used in the exemplification of the present invention.

Sequence 386 AA;  
Query Match 100.0%; Score 61; DB 9; Length 386;  
Best Local Similarity 100.0%; Pred. No. 0.06;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10  
| | | | | | | | | |  
DB 280 CTNTDNPYK 289

RESULT 13  
ADV27061  
ID ADV27061 standard; protein; 460 AA.

AC ADV27061;  
XX 05-MAY-2005 (first entry)  
XX Heparanase inhibitor protein #4.  
XX Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
KW neurological disease; viral infection; infection; cytostatic;  
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
KW heparanase modulator; enzyme purification.

XX Homo sapiens.  
OS  
XX WO2005016227-A2.  
PN  
XX 24-FEB-2005.  
PD  
XX 12-AUG-2004; 2004WO-IL000744.  
PF  
XX 14-AUG-2003; 2003US-0494800P.  
PR  
XX 12-JAN-2004; 2004US-0535492P.  
PR  
XX (INSI-) INSTIGHT BIOPHARMACEUTICALS LTD.

PA Van-Gelder JM, Miron D;  
XX WPI; 2005-182203/19.  
XX  
XX Regulating heparanase activity, useful for treating heparanase-associated  
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
PT neurological diseases or viral diseases) comprises modulating heparanase  
PT activation.

XX Disclosure; SEQ ID NO 37; 211pp; English.  
PS  
XX The invention relates to a method of regulating heparanase activity in a  
CC tissue or regulating a biological process depending at least in part on

CC heparanase activity comprising modulating heparanase activation. The  
CC invention also relates to methods of treating a heparanase- or heparin  
CC binding protein-associated disease or disorder in a subject, a  
CC pharmaceutical composition for use in the treatment of a heparanase-  
CC associated disease or disorder comprising a therapeutic amount of an  
CC agent capable of modulating heparanase activation and a pharmaceutical  
CC carrier or diluent, a method of identifying a protease activator of  
CC heparanase, a protease substrate mimetic comprising a peptide  
CC representing a subset or all substrate residues or cleavage sites of  
CC human heparanase or an equivalent non-human heparanase, a method of  
CC producing active heparanase and a method of modulating an adhesion  
CC activity of heparanase. The composition and methods are useful for  
CC modulating heparanase activation and for treating heparanase-associated  
CC diseases or disorders such as cancer, inflammation, cardiovascular  
CC diseases, neurological diseases or viral infections. This sequence  
CC represents a heparanase inhibitor protein used in the scope of the  
CC invention.

XX Sequence 460 AA;

Query Match 100.0%; Score 61; DB 9; Length 460;  
Best Local Similarity 100.0%; Pred. No. 0.071;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10  
| | | | | | | | | |  
DB 354 CTNTDNPYK 363

RESULT 14  
AEB87589  
ID AEB87589 standard; protein; 486 AA.

AC AEB87589;  
XX 06-OCT-2005 (first entry)  
XX Human heparanase 65delta20 deletion mutant.  
XX Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;  
KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;  
KW leukemia; lymphoma; dermatological disease; hematological disease;  
KW immune disorder; inflammation; antiinflammatory; renal disease;  
KW nephrotropic; endocrine disease; genitourinary disease;  
KW autoimmune disease; immunosuppressive; drug screening; muten.

OS Homo sapiens.  
OS Synthetic.

PN WO2005071070-A2.

XX 04-AUG-2005.

PD 20-JAN-2005; 2005WO-IL0000068.

PF 22-JAN-2004; 2004IL-00160025.

PR 28-JUL-2004; 2004US-00901943.

PR (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX Vlodavsky I, Ilan N, Levy-Adam F;

XX WPI; 2005-564219/57.

DR N-FSDB; AEB87588.

XX New amino acid sequences derived from the 50 kDa subunit of heparanase,  
PT for treating or inhibiting malignant proliferative, inflammatory, kidney  
PT disorder or autoimmune disorder.

XX Claim 107; SEQ ID NO 31; 167pp; English.

XX The present sequence is that of a deletion mutant of human heparanase,  
CC denoted 65delta20, which is devoid of amino acid residues 411-432 of the

CC native protein. The recombinant protein is deficient of heparanase  
 CC endoglycosidase catalytic activity. The invention relates to amino acid  
 CC sequences derived from the N-terminus region of the 50 kDa subunit of  
 CC heparanase, particularly in the regions between amino acid residues 158-  
 CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise  
 CC heparin-binding domains. The invention also provides an antibody directed  
 CC to these sequences, in particular the 158-171 peptide, and compositions  
 CC and uses of this antibody as a heparanase inhibitor. A screening method  
 CC is provided for specific heparanase inhibitors. Claimed pharmaceutical  
 CC compositions comprising (i) a peptide derived from the N-terminus region  
 CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a  
 CC peptide, or (iii) an antibody which specifically recognizes the peptide  
 CC are used for the inhibition of heparanase catalytic activity associated  
 CC with an inflammatory disorder, kidney disease, autoimmune disease,  
 CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or  
 CC with a malignant proliferative disorder, especially a solid or non-solid  
 CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.  
 XX  
 SQ Sequence 486 AA;

Query Match 100.0%; Score 61; DB 9; Length 486;  
 Best Local Similarity 100.0%; Pred. No. 0.075;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10  
 |||||  
 DB 380 CTNTDNPYK 389

RESULT 15  
 ADZ18996  
 ID ADZ18996 standard; protein; 492 AA.  
 AC ADZ18996;  
 XX  
 DT 16-JUN-2005 (first entry)  
 XX  
 DE Hep106 construct protein.  
 XX  
 KW Enzyme engineering; heparanase; hepi106; metastasis; autoimmune disease;  
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
 KW immunosuppressive; enzyme.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2005030962-A1.  
 XX  
 PD 07-APR-2005.  
 XX  
 PF 17-SEP-2004; 2004WO-EP010517.  
 XX  
 PR 26-SEP-2003; 2003US-0506479P.  
 PR 20-JAN-2004; 2004US-0537729P.  
 XX  
 PA (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.  
 XX  
 PI Lahm A, Nardella C, Pallaooro M, Steinkuhler C;  
 XX  
 DR WPI; 2005-273382/28.  
 DR N-PSDB; ADZ18997.  
 XX  
 PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.  
 XX  
 PS Example 2; SEQ ID NO 17; 65pp; English.  
 XX  
 CC The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoproteinase. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in  
 CC inhibitor screening assays for the development of therapeutics or

CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a hepi106 construct protein  
 CC used in the scope of the invention.  
 XX  
 SQ Sequence 492 AA;

Query Match 100.0%; Score 61; DB 9; Length 492;  
 Best Local Similarity 100.0%; Pred. No. 0.076; 0; Indels 0; Gaps 0;  
 Matches 10; Conservative 0; Mismatches 0;  
 QY 1 CTNTDNPYK 10  
 |||||  
 DB 386 CTNTDNPYK 395

RESULT 16  
 AEB87562  
 ID AEB87562 standard; protein; 493 AA.  
 AC AEB87562;  
 XX  
 DT 06-OCT-2005 (first entry)  
 XX  
 DE Human heparanase 65delta15 deletion mutant.  
 XX  
 KW Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;  
 KW cytostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;  
 KW leukemia; lymphoma; dermatological disease; hematological disease;  
 KW immune disorder; inflammation; antiinflammatory; renal disease;  
 KW nephrotropic; endocrine disease; genitourinary disease;  
 KW autoimmune disease; immunosuppressive; drug screening; mutein.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO2005071070-A2.  
 XX  
 PD 04-AUG-2005.  
 XX  
 PF 20-JAN-2005; 2005WO-IL000068.  
 XX  
 PR 22-JAN-2004; 2004IL-00160025.  
 PR 28-JUL-2004; 2004US-00901943.  
 XX  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.  
 XX  
 PI Vlodavsky I, Ilan N, Levy-Adam F;  
 XX  
 DR WPI; 2005-564219/57.  
 DR N-PSDB; AEB87561.  
 XX  
 PT New amino acid sequences derived from the 50 kDa subunit of heparanase,  
 PT for treating or inhibiting malignant proliferative, inflammatory, kidney  
 PT disorder or autoimmune disorder.  
 XX  
 PS Claim 105; SEQ ID NO 4; 167pp; English.  
 XX  
 CC The present sequence is that of a deletion mutant of human heparanase,  
 CC denoted 65delta15, which is devoid of amino acid residues 158-171 of the  
 CC native protein. The recombinant protein is deficient of heparanase  
 CC endoglycosidase catalytic activity. The invention relates to amino acid  
 CC sequences derived from the N-terminus region of the 50 kDa subunit of  
 CC heparanase, particularly in the regions between amino acid residues 158-  
 CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise  
 CC heparin-binding domains. The invention also provides an antibody directed  
 CC to these sequences, in particular the 158-171 peptide, and compositions  
 CC and uses of this antibody as a heparanase inhibitor. A screening method  
 CC is provided for specific heparanase inhibitors. Claimed pharmaceutical  
 CC compositions comprising (i) a peptide derived from the N-terminus region  
 CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a  
 CC peptide, or (iii) an antibody which specifically recognizes the peptide  
 CC are used for the inhibition of heparanase catalytic activity associated  
 CC with an inflammatory disorder, kidney disease, autoimmune disease,

CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or  
 CC with a malignant proliferative disorder, especially a solid or non-solid  
 CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.  
 XX  
 SQ Sequence 493 AA;

Query Match 100.0%; Score 61; DB 9; Length 493;  
 Best Local Similarity 100.0%; Pred. No. 0.076;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10  
 Db 387 CTNTDNPYK 396  
 |||||

RESULT 17  
 ADZ18999  
 ID ADZ18999 standard; protein; 495 AA.

XX AC ADZ18999;

DT 16-JUN-2005 (first entry)

DE Hep109 construct protein.

XX Enzyme engineering; heparanase; hepl09; metastasis; autoimmune disease;  
 KW inflammation; neoplasm; immune disorder; antiinflammatory; cystostatic;  
 KW immunosuppressive; enzyme.

XX Synthetic.

XX WO2005030962-A1.

XX 07-APR-2005.

XX 17-SEP-2004; 2004WO-EP010517.

XX 26-SEP-2003; 2003US-0506479P.

XX 20-JAN-2004; 2004US-0537729P.

XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.

XX Lahm A, Nardella C, Pallaoro M, Steinkuhler C;

XX WPI; 2005-273382/28.

XX N-PSDB; ADZ18998.

XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.

XX Example 2; SEQ ID NO 20; 65pp; English.

XX The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoproteinase. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in  
 CC inhibitor screening assays for the development of therapeutics or  
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a hepl09 construct protein  
 CC used in the scope of the invention.

XX Sequence 495 AA;

Query Match 100.0%; Score 61; DB 9; Length 495;  
 Best Local Similarity 100.0%; Pred. No. 0.077;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10  
 Db 389 CTNTDNPYK 398  
 |||||

RESULT 18  
 AEB87587

ID AEB87587 standard; protein; 497 AA.

XX AC AEB87587;

XX 06-OCT-2005 (first entry)

XX Human heparanase 65delta10 deletion mutant.

XX Heparanase inhibitor; angiogenesis inhibition; tumor; neoplasm;  
 KW cystostatic; metastasis; hyperproliferation; carcinoma; sarcoma; melanoma;  
 KW leukemia; lymphoma; dermatological disease; hematological disease;  
 KW immune disorder; inflammation; antiinflammatory; renal disease;  
 KW nephrotropic; endocrine disease; genitourinary disease;  
 KW autoimmune disease; immunosuppressive; drug screening; mutein.

XX Homo sapiens.

XX Synthetic.

XX WO2005071070-A2.

XX 04-AUG-2005.

XX 20-JAN-2005; 2005WO-IL000068.

XX 22-JAN-2004; 2004IL-00160025.

XX 28-JUL-2004; 2004US-00901943.

XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX Vlodayvsky I, Ilan N, Levy-Adam F;

XX WPI; 2005-564219/57.

XX N-PSDB; AEB87586.

XX New amino acid sequences derived from the 50 kDa subunit of heparanase,  
 PT for treating or inhibiting malignant proliferative, inflammatory, kidney  
 PT disorder or autoimmune disorder.

XX Claim 106; SEQ ID NO 29; 167pp; English.

XX The present sequence is that of a deletion mutant of human heparanase,  
 CC denoted 65delta10, which is devoid of amino acid residues 270-280 of the  
 CC native protein. The recombinant protein is deficient of heparanase  
 CC endoglycosidase catalytic activity. The invention relates to amino acid  
 CC sequences derived from the N-terminus region of the 50 kDa subunit of  
 CC heparanase, particularly in the regions between amino acid residues 158-  
 CC 171, 270-280 and 411-432 of human heparanase. These sequences comprise  
 CC heparin-binding domains. The invention also provides an antibody directed  
 CC to these sequences, in particular the 158-171 peptide, and compositions  
 CC and uses of this antibody as a heparanase inhibitor. A screening method  
 CC is provided for specific heparanase inhibitors. Claimed pharmaceutical  
 CC compositions comprising (i) a peptide derived from the N-terminus region  
 CC of the heparanase 50 kDa subunit, (ii) a substance which binds to such a  
 CC peptide, or (iii) an antibody which specifically recognizes the peptide  
 CC are used for the inhibition of heparanase catalytic activity associated  
 CC with an inflammatory disorder, kidney disease, autoimmune disease,  
 CC angiogenesis, tumor formation, tumor progression or tumor metastasis, or  
 CC with a malignant proliferative disorder, especially a solid or non-solid  
 CC tumor selected from a carcinoma, sarcoma, melanoma, leukemia or lymphoma.

XX Sequence 497 AA;

Query Match 100.0%; Score 61; DB 9; Length 497;  
 Best Local Similarity 100.0%; Pred. No. 0.077;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10  
 Db 391 CTNTDNPYK 400  
 |||||

```

RESULT 19
ADZ19000
ID  ADZ19000 standard; protein; 501 AA.
XX
XX  ADZ19000;
AC
XX  16-JUN-2005 (first entry)
DT
XX
DE  HepG3 construct protein.
XX
KW  Enzyme engineering; heparanase; hepgS3; metastasis; autoimmune disease;
KW  inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW  immunosuppressive; enzyme.
XX
XX  Synthetic.
OS
XX
XX  WO2005030962-A1.
PN
XX
XX  07-APR-2005.
PD
XX
XX  17-SEP-2004; 2004WO-EP010517.
PF
XX
XX  26-SEP-2003; 2003US-0506479P.
PR
XX  20-JAN-2004; 2004US-0537729P.
PR
XX
XX  (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
PA
XX
XX  Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
PI
XX
XX  WPI; 2005-273382/28.
DR
XX  N-PSDB; ADZ19001.
DR
XX
XX  Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT  nucleotide sequence that encodes mammalian heparanase protein and has two
PT  consensus cleavage sites located between specific nucleotide encoding
PT  residues.
XX
XX  Example 2; SEQ ID NO 21; 65pp; English.
PS
XX
XX  The invention relates to a synthetic nucleic acid molecule that encodes
CC  mammalian heparanase protein, where the nucleic acid comprises two
CC  consensus cleavage sites recognized by endoproteinase. The sequences are
CC  useful for expressing mammalian heparanase in non-mammalian cells and in
CC  inhibitor screening assays for the development of therapeutics or
CC  pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC  and/or inflammation. This sequence represents a hepgS3 construct protein
CC  used in the scope of the invention.
XX
XX  Sequence 501 AA;
SQ
XX
Query Match      100.0%; Score 61; DB 9; Length 501;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTNTDNPYK 10
Db      |||||
        395 CTNTDNPYK 404

RESULT 20
ADZ19005
ID  ADZ19005 standard; protein; 507 AA.
XX
XX  ADZ19005;
AC
XX  16-JUN-2005 (first entry)
DT
XX
XX  HepG6 construct protein.
DE
XX
KW  Enzyme engineering; heparanase; hepgS6; metastasis; autoimmune disease;
KW  inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW  immunosuppressive; enzyme.
XX

```

```

XX  Synthetic.
OS
XX  WO2005030962-A1.
PN
XX
XX  07-APR-2005.
PD
XX
XX  17-SEP-2004; 2004WO-EP010517.
PF
XX
XX  26-SEP-2003; 2003US-0506479P.
PR
XX  20-JAN-2004; 2004US-0537729P.
PR
XX
XX  (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.
PA
XX
XX  Lahm A, Nardella C, Pallaoro M, Steinkuhler C;
PI
XX
XX  WPI; 2005-273382/28.
DR
XX  N-PSDB; ADZ19003.
DR
XX
XX  Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT  nucleotide sequence that encodes mammalian heparanase protein and has two
PT  consensus cleavage sites located between specific nucleotide encoding
PT  residues.
XX
XX  Example 2; SEQ ID NO 26; 65pp; English.
PS
XX
XX  The invention relates to a synthetic nucleic acid molecule that encodes
CC  mammalian heparanase protein, where the nucleic acid comprises two
CC  consensus cleavage sites recognized by endoproteinase. The sequences are
CC  useful for expressing mammalian heparanase in non-mammalian cells and in
CC  inhibitor screening assays for the development of therapeutics or
CC  pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC  and/or inflammation. This sequence represents a hepgS6 construct protein
CC  used in the scope of the invention.
XX
XX  Sequence 507 AA;
SQ
XX
Query Match      100.0%; Score 61; DB 9; Length 507;
Best Local Similarity 100.0%; Pred. No. 0.079;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CTNTDNPYK 10
Db      |||||
        401 CTNTDNPYK 410

RESULT 21
ADY27058
ID  ADY27058 standard; protein; 508 AA.
XX
XX  ADY27058;
AC
XX
XX  05-MAY-2005 (first entry)
DT
XX
XX  Human inactive heparanase protein.
DE
XX
XX  Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
KW  neurological disease; viral infection; infection; cytostatic;
KW  antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
KW  protease; enzyme; enzyme purification.
XX
XX  Homo sapiens.
OS
XX
XX  WO2005016227-A2.
PN
XX
XX  24-FEB-2005.
PD
XX
XX  12-AUG-2004; 2004WO-IL000744.
PF
XX
XX  14-AUG-2003; 2003US-0494800P.
PR
XX  12-JAN-2004; 2004US-0535492P.
PR
XX
XX  (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
PA

```

XX PI Van-Gelder JM, Miron D;  
 XX WPI; 2005-182203/19.  
 XX  
 XX Regulating heparanase activity, useful for treating heparanase-associated  
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
 PT neurological diseases or viral diseases) comprises modulating heparanase  
 PT activation.  
 XX  
 XX Claim 257; SEQ ID NO 34; 211pp; English.  
 XX  
 XX The invention relates to a method of regulating heparanase activity in a  
 CC tissue or regulating a biological process depending at least in part on  
 CC heparanase activity comprising modulating heparanase activation. The  
 CC invention also relates to methods of treating a heparanase- or heparin  
 CC binding protein-associated disease or disorder in a subject, a  
 CC pharmaceutical composition for use in the treatment of a heparanase-  
 CC associated disease or disorder comprising a therapeutic amount of an  
 CC agent capable of modulating heparanase activation and a pharmaceutical  
 CC carrier or diluent, a method of identifying a protease activator of  
 CC heparanase, a protease substrate mimetic comprising a peptide  
 CC representing a subset or all substrate residues or cleavage sites of  
 CC human heparanase or an equivalent non-human heparanase, a method of  
 CC producing active heparanase and a method of modulating an adhesion  
 CC activity of heparanase. The composition and methods are useful for  
 CC modulating heparanase activation and for treating heparanase-associated  
 CC diseases or disorders such as cancer, inflammation, cardiovascular  
 CC diseases, neurological diseases or viral infections. This sequence  
 CC represents a human inactive heparanase protein used in the scope of the  
 CC invention.  
 XX  
 XX Sequence 508 AA;  
 SQ  
 Query Match 100.0%; Score 61; DB 9; Length 508;  
 Best Local Similarity 100.0%; Pred. No. 0.079;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 CTNTDNPYK 10  
 Db 402 CTNTDNPYK 411  
 |||||  
 RESULT 22  
 ADZ19006  
 ID ADZ19006 standard; protein; 526 AA.  
 XX  
 XX AC ADZ19006;  
 XX  
 XX 16-JUN-2005 (first entry)  
 DT  
 XX Hephyluro construct protein.  
 DE  
 XX Enzyme engineering; heparanase; hephyluro; metastasis;  
 KW autoimmune disease; inflammation; neoplasm; immune disorder;  
 KW antiinflammatory; cytostatic; immunosuppressive; enzyme.  
 XX  
 XX Synthetic.  
 OS  
 XX WO2005030962-A1.  
 PN  
 XX 07-APR-2005.  
 PD  
 XX 17-SEP-2004; 2004WO-EP010517.  
 PF  
 XX 26-SEP-2003; 2003US-0506479P.  
 PR  
 XX 20-JAN-2004; 2004US-0537729P.  
 PR  
 XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.  
 PA Lahm A, Nardella C, Pallao M, Steinkuhler C;  
 PI WPI; 2005-273382/28.  
 XX  
 XX

DR N-PSDB; ADZ19007.  
 XX  
 XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
 PT nucleotide sequence that encodes mammalian heparanase protein and has two  
 PT consensus cleavage sites located between specific nucleotide encoding  
 PT residues.  
 XX  
 XX Example 2; SEQ ID NO 27; 65pp; English.  
 XX  
 XX The invention relates to a synthetic nucleic acid molecule that encodes  
 CC mammalian heparanase protein, where the nucleic acid comprises two  
 CC consensus cleavage sites recognized by endoproteinase. The sequences are  
 CC useful for expressing mammalian heparanase in non-mammalian cells and in  
 CC inhibitor screening assays for the development of therapeutics or  
 CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
 CC and/or inflammation. This sequence represents a hephyluro construct  
 CC protein used in the scope of the invention.  
 XX  
 XX Sequence 526 AA;  
 SQ  
 Query Match 100.0%; Score 61; DB 9; Length 526;  
 Best Local Similarity 100.0%; Pred. No. 0.081;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 CTNTDNPYK 10  
 Db 420 CTNTDNPYK 429  
 |||||  
 RESULT 23  
 ABB07815  
 ID ABB07815 standard; protein; 527 AA.  
 XX  
 XX AC ABB07815;  
 XX  
 XX 03-JUL-2002 (first entry)  
 DT  
 XX Chicken signal peptide/human heparanase chimeric protein sequence.  
 DE  
 XX Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;  
 KW anti-protozoan; neuroprotective; heparin; chicken; human; chimeric.  
 KW  
 XX Synthetic.  
 OS  
 XX Gallus gallus.  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT Peptide i. 19  
 FT Protein /note= "chicken heparanase signal peptide"  
 FT 20. 527  
 FT /note= "human heparanase mature protein"  
 XX  
 XX US2002034810-A1.  
 PN  
 XX 21-MAR-2002.  
 PD  
 XX 16-AUG-2001; 2001US-00930218.  
 PF  
 XX 20-SEP-2000; 2000US-00666390.  
 PR  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA  
 XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia B;  
 PI WPI; 2002-338926/37.  
 DR  
 XX N-PSDB; ABL40753.  
 XX  
 XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful  
 PT to treat various heparin-related disorders and the signal peptide is  
 PT useful in production of membrane-targeted or secreted recombinant  
 PT proteins.  
 XX  
 XX Disclosure; Page 26-28; 39pp; English.  
 PS

XX The invention relates to an isolated avian and reptile nucleic acid,  
 CC encoding a polypeptide with heparanase catalytic activity. The signal  
 CC peptide of the nucleic acid can be used to express membrane-associated or  
 CC secreted proteins in heterologous expression systems. The encoded  
 CC polypeptides can be used to prevent tumour angiogenesis, metastasis and  
 CC invasion, and to intervene with pathologies associated with impaired  
 CC heparin-binding growth factors, cellular responses to heparin-binding  
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,  
 CC cellular susceptibility to viral, protozoa and bacterial infections or  
 CC disintegration of neurodegenerative plaques. The present sequence  
 CC represents a chicken signal peptide/human heparanase chimeric protein  
 XX  
 SQ Sequence 527 AA;

Query Match 100.0%; Score 61; DB 5; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 0.082;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTNTDNPYK 10  
 Db 421 CTNTDNPYK 430  
 |||||

RESULT 24  
 ABW02018  
 ID ABW02018 standard; protein; 527 AA.  
 XX  
 AC ABW02018;  
 XX  
 DT 12-FEB-2004 (first entry)  
 XX  
 DE Chimeric human-chicken heparanase protein.  
 XX  
 KW Chicken; heparanase; tumour cell metastasis; inflammation; autoimmunity;  
 KW wound healing; angiogenesis; restenosis; Genstmann-Straussler Syndrome;  
 KW neurodegenerative disease; atherosclerosis; Creutzfeldt-Jakob disease;  
 KW infection; Scrapie; Alzheimer's disease; protein therapy; cytostatic;  
 KW immunosuppressive; vulnery; bactericide; anti-angiogenic; virucide;  
 KW antisclerotic; neuroprotective; protozoacide; chimeric; fusion protein;  
 KW enzyme; human.  
 XX  
 OS Chimeric - Gallus gallus.  
 OS Chimeric - Homo sapiens.  
 XX  
 PN US2003180788-A1.  
 XX  
 PD 25-SEP-2003.  
 XX  
 PF 08-MAY-2003; 2003US-00431438.  
 XX  
 PR 20-SEP-2000; 2000US-00666390.  
 PR 16-AUG-2001; 2001US-00930218.  
 XX  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX  
 PI Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;  
 XX  
 DR WPI; 2003-843931/78.  
 DR N-PSDB; AAD63532.  
 XX  
 FT Recombinant jungle red fowl (Gallus gallus) heparanase protein, useful  
 FT for treating cancers, microbial infections and aiding wound healing.  
 XX  
 PS Example; Page 26-28; Opp; English.  
 XX  
 CC The present invention relates to novel jungle red fowl heparanase protein  
 CC and polynucleotides encoding such proteins. Heparanase sequences can be  
 CC used to develop treatments for various diseases, to develop diagnostic  
 CC assays for these diseases and to provide new tools for basic and directed  
 CC research especially in the fields of medicine and biology. They can be

CC used to develop new drugs to inhibit tumour cell metastasis, inflammation  
 CC and autoimmunity. Recombinant heparanase offers a potential treatment for  
 CC wound healing, angiogenesis, restenosis, atherosclerosis, inflammation,  
 CC neurodegenerative diseases (e.g. Genstmann-Straussler Syndrome, Scrapie,  
 CC Creutzfeldt-Jakob disease and Alzheimer's disease) and certain viral and  
 CC some bacterial and protozoa infections. Recombinant heparanase can also  
 CC be used to neutralise plasma heparin, as a potential replacement of  
 CC protamine. Sequences of the invention are used in protein therapy. The  
 CC present sequence is chimeric human-chicken heparanase protein  
 XX  
 SQ Sequence 527 AA;

Query Match 100.0%; Score 61; DB 7; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 0.082;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTNTDNPYK 10  
 Db 421 CTNTDNPYK 430  
 |||||

RESULT 25  
 ADO63825  
 ID ADO63825 standard; protein; 527 AA.  
 XX  
 AC ADO63825;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Chimeric heparanase mutant E225A, SEQ ID:10.  
 XX  
 KW Human; chicken; heparanase; heparanase-derived protein;  
 KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;  
 KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;  
 KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnerary; mutant; mutein.  
 XX  
 OS Homo sapiens.  
 OS Gallus gallus.  
 OS Synthetic.  
 OS Chimeric.  
 XX  
 XX Key Location/Qualifiers  
 FT Peptide 1..18  
 FT /note= "Chicken heparanase signal peptide"  
 FT Region 19..527  
 FT /note= "Corresponds to residues 35-543 of human  
 FT heparanase mutant E225A (SEQ ID NO:7)"  
 FT Misc-difference 209  
 FT /note= "Ala replaces wild-type Glu (active site proton  
 FT donor). Corresponds to residue 225 of human heparanase  
 FT mutant E225A (SEQ ID NO:7)"  
 FT Active-site 327  
 FT /note= "Active site nucleophile"  
 XX  
 PN WO2004048558-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 24-NOV-2003; 2003WO-IL000989.  
 XX  
 PR 24-NOV-2002; 2002IL-00153059.  
 XX  
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA  
 XX Vlodavsky I, Zcharia E, Goldshmidt O, Ilan N;  
 XX  
 DR WPI; 2004-450373/42.  
 DR N-PSDB; ADO63819.  
 XX  
 PT New nucleic acid construct comprising heparanase-derived polypeptide,

PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 XX thrombasthenia, or Bernard-Soulier syndrome.  
 PS Claim 10; SEQ ID NO 10; 128pp; English.  
 XX  
 CC The invention relates to nucleic acid constructs comprising a nucleic  
 CC acid encoding a heparanase-derived protein which lacks heparanase  
 CC endoglycosidase catalytic activity but which retains its cell-cell and  
 CC cell-matrix adhesion properties. The constructs of the invention  
 CC optionally further comprise operably linked regulatory elements. The  
 CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the  
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
 CC host cells are useful in preparing a tissue sealant composition for  
 CC sealing injuries, reducing the loss of blood, accelerating the healing  
 CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialisation of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be used in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte  
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents a chimeric  
 CC protein comprising the signal peptide of chicken heparanase and residues  
 CC 35-543 of the human heparanase mutant E225A.  
 XX  
 SQ Sequence 527 AA;

Query Match 100.0%; Score 61; DB 8; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 0.082;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10  
 |||||  
 Db 421 CTNTDNPYK 430

RESULT 26  
 ADO63827  
 ID ADO63827 standard; protein; 527 AA.  
 AC ADO63827;  
 XX  
 XX 26-AUG-2004 (first entry)  
 XX  
 DE Chimeric heparanase mutant E225A/E343A, SEQ ID:12.  
 XX  
 KW Human; chicken; heparanase; heparanase-derived protein;  
 KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;  
 KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;  
 KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnary; mutant; mutein.  
 XX  
 OS Homo sapiens.  
 OS Gallus gallus.  
 OS Synthetic.  
 OS Chimeric.  
 XX  
 XX Key Location/Qualifiers  
 FH Peptide 1..18  
 FT /note= "Chicken heparanase signal peptide"  
 FT 19..527  
 FT Region  
 FT /note= "Corresponds to residues 35-543 of human  
 FT heparanase mutant E225A/E343A (SEQ ID NO:9)"  
 FT

Misc-difference 209  
 /note= "Ala replaces wild-type Glu (active site proton  
 donor). Corresponds to residue 225 of human heparanase  
 mutant E225A/E343A (SEQ ID NO:9)"  
 FT  
 FT  
 FT Misc-difference 327  
 /note= "Ala replaces wild-type Glu (active site  
 nucleophile). Corresponds to residue 343 of human  
 heparanase mutant E225A/E343A (SEQ ID NO:9)"  
 FT  
 FT  
 PN WO2004048558-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 XX 24-NOV-2003; 2003WO-IL000989.  
 PF  
 XX 24-NOV-2002; 2002IL-00153059.  
 PR  
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA  
 XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
 PI  
 XX WPI; 2004-450373/42.  
 DR  
 DR N-PSDB; ADO63821.  
 XX  
 FT New nucleic acid construct comprising heparanase-derived polypeptide,  
 FT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 FT thrombasthenia, or Bernard-Soulier syndrome.  
 FT  
 XX Claim 10; SEQ ID NO 12; 128pp; English.  
 PS  
 XX The invention relates to nucleic acid constructs comprising a nucleic  
 CC acid encoding a heparanase-derived protein which lacks heparanase  
 CC endoglycosidase catalytic activity but which retains its cell-cell and  
 CC cell-matrix adhesion properties. The constructs of the invention  
 CC optionally further comprise operably linked regulatory elements. The  
 CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the  
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
 CC host cells are useful in preparing a tissue sealant composition for  
 CC sealing injuries, reducing the loss of blood, accelerating the healing  
 CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialisation of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be used in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte  
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents a chimeric  
 CC protein comprising the signal peptide of chicken heparanase and residues  
 CC 35-543 of the human heparanase double mutant E225A/E343A.  
 XX  
 SQ Sequence 527 AA;

Query Match 100.0%; Score 61; DB 8; Length 527;  
 Best Local Similarity 100.0%; Pred. No. 0.082;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10  
 |||||  
 Db 421 CTNTDNPYK 430

RESULT 27  
 ADO63826  
 ID ADO63826 standard; protein; 527 AA.  
 XX



AC ADO63826;  
XX 26-AUG-2004 (first entry)  
XX Chimeric heparanase mutant E343A, SEQ ID:11.  
XX Human; chicken; heparanase; heparanase-derived protein;  
KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;  
KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;  
KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;  
KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
KW vulnary; mutant; mutein.  
XX Homo sapiens.  
OS Gallus gallus.  
OS Synthetic.  
OS Chimeric.  
XX Key Location/Qualifiers  
XX Peptide 1..18  
XX /note= "Chicken heparanase signal peptide"  
XX Region 19..527  
XX /note= "Corresponds to residues 35-543 of human  
XX heparanase mutant E343A (SEQ ID NO:8)"  
XX Active-site 209  
XX /note= "Active site proton donor"  
XX Misc-difference 327  
XX /note= "Ala replaces wild-type Glu (active site  
XX nucleophile). Corresponds to residue 343 of human  
XX heparanase mutant E343A (SEQ ID NO:8)"  
XX PN W02004048558-A2.  
XX 10-JUN-2004.  
XX 24-NOV-2003; 2003WO-IL000989.  
XX 24-NOV-2002; 2002IL-00153059.  
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
XX WPI; 2004-450373/42.  
XX N-PSDB; ADO63820.  
XX New nucleic acid construct comprising heparanase-derived polypeptide,  
XX useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
XX thrombasthenia, or Bernard-Soulier syndrome.  
XX Claim 10; SEQ ID NO 11, 128pp; English.  
XX The invention relates to nucleic acid constructs comprising a nucleic  
XX acid encoding a heparanase-derived protein which lacks heparanase  
XX endoglycosidase catalytic activity but which retains its cell-cell and  
XX cell-matrix adhesion properties. The constructs of the invention  
XX optionally further comprise operably linked regulatory elements. The  
XX invention also relates to the heparanase-derived proteins and host cells  
XX comprising the nucleic acid constructs of the invention. The heparanase-  
XX derived proteins are especially mutants of human heparanase in which the  
XX active site proton donor Glu225 and/or the active site nucleophile Glu343  
XX are replaced with Ala (ADO63822-ADO63824), and the proteins may  
XX optionally further comprise an avian heparanase signal peptide (ADO63825-  
XX ADO63827). The heparanase-derived protein, nucleic acid construct and  
XX host cells are useful in preparing a tissue sealant composition for  
XX sealing injuries, reducing the loss of blood, accelerating the healing  
XX and homeostasis of an injury, accelerating blood vessel endothelium  
XX formation or the endothelialisation of vascular grafts, accelerating the  
XX adhesive activity of mammalian cells, and accelerating the adhesion and  
XX aggregation of platelets. They may also be use in the treatment of  
XX disorders associated with adhesion deficiency such as LAD (leukocyte  
XX adhesion deficiency), Glanzmann's thrombasthenia (defective platelet

CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
CC cells of the invention may additionally be to screen for modulators of  
CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
CC synthetic vascular graft comprising a tube made of a biocompatible  
CC material lined with the cells. The present sequence represents a chimeric  
CC protein comprising the signal peptide of chicken heparanase and residues  
CC 35-543 of the human heparanase mutant E343A.  
XX Sequence 527 AA;  
SQ Query Match 100.0%; Score 61; DB 8; Length 527;  
Best Local Similarity 100.0%; Pred. No. 0.082;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CTNTDNPYK 10  
DB 421 CTNTDNPYK 430  
RESULT 28  
ADZ19004  
ID ADZ19004 standard; protein; 527 AA.  
XX AC ADZ19004;  
XX DT 16-JUN-2005 (first entry)  
XX DE HepGS4 construct protein.  
XX KW Enzyme engineering; heparanase; hepGS4; metastasis; autoimmune disease;  
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
KW immunosuppressive; enzyme.  
XX OS Synthetic.  
XX PN W02005030962-A1.  
XX PD 07-APR-2005.  
XX PF 17-SEP-2004; 2004WO-EP010517.  
XX PR 26-SEP-2003; 2003US-0506479P.  
XX PR 20-JAN-2004; 2004US-0537729P.  
XX PA (RICE-) IST RICERCH BIOL MOLECOLARE ANGELETTI.  
XX PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;  
XX WPI; 2005-273382/28.  
XX DR N-PSDB; ADZ19002.  
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
XX nucleotide sequence that encodes mammalian heparanase protein and has two  
XX consensus cleavage sites located between specific nucleotide encoding  
XX residues.  
XX Example 2; SEQ ID NO 25; 65pp; English.  
XX The invention relates to a synthetic nucleic acid molecule that encodes  
XX mammalian heparanase protein, where the nucleic acid comprises two  
XX consensus cleavage sites recognized by endoproteinase. The sequences are  
XX useful for expressing mammalian heparanase in non-mammalian cells and in  
XX inhibitor screening assays for the development of therapeutics or  
XX pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
XX and/or inflammation. This sequence represents a hepGS4 construct protein  
XX used in the scope of the invention.  
SQ Sequence 527 AA;  
Query Match 100.0%; Score 61; DB 9; Length 527;  
Best Local Similarity 100.0%; Pred. No. 0.082;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



XX DE A human heparanase protein.  
 XX KW Heparanase; hp; modulator; heparin-binding growth factor;  
 XX cellular response; cytokine; cell interaction; plasma lipoprotein;  
 KW cellular susceptibility; infection; disintegration;  
 KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;  
 KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;  
 KW plasma heparin; micrometastasis; autoimmune lesion; renal failure.  
 XX OS Homo sapiens.  
 XX KW WO9911798-A1.  
 XX PD 11-MAR-1999.  
 XX KW 31-AUG-1998; 98WO-US017954.  
 XX PF 02-SEP-1997; 97US-00922170.  
 XX PR 02-JUL-1998; 98US-00109386.  
 XX KW (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX PA (FRIE/) FRIEDMAN M M.  
 XX PI Pecker I, Vlodavsky I, Feinstein E;  
 XX WPI; 1999-302255/25.  
 XX DR N-PSDB; AAX35648.  
 XX KW New human polynucleotide useful for treating angiogenesis, restenosis,  
 PT and inflammation.  
 XX PS Claim 6; Fig 1; 63pp; English.  
 XX CC The specification describes a polypeptide having heparanase (hp)  
 CC activity. The recombinant protein is used as a modulator of heparin-  
 CC binding growth factors, cellular responses to heparin-binding growth  
 CC factors and cytokines, cell interaction with plasma lipoproteins,  
 CC cellular susceptibility to viral, protozoal and bacterial infections or  
 CC disintegration of neurodegenerative plaques. Heparanase may be useful for  
 CC conditions such as wound healing, angiogenesis, restenosis,  
 CC atherosclerosis, inflammation, neurodegenerative diseases, and viral  
 CC infections. Mammalian heparanase can be used to neutralize plasma  
 CC heparin, and anti-heparanase antibodies may be applied for  
 CC immunodetection and diagnosis of micrometastases, autoimmune lesions, and  
 CC renal failure in biopsy specimens, plasma samples, and body fluids. The  
 CC present sequence represents human heparanase  
 XX Sequence 543 AA;  
 SQ

Query Match 100.0%; Score 61; DB 2; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 0.084;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTNTDNPYK 10  
 |||||  
 Db 437 CTNTDNPYK 446

RESULT 32  
 AAY17082  
 ID AAY17082 standard; protein; 543 AA.  
 XX AC AAY17082;  
 XX DT 21-JUL-1999 (first entry)  
 XX DE Human heparanase enzyme.  
 XX KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;  
 KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;  
 KW arteriosclerosis; atherosclerosis; inflammation; tissue development;

KW human; HSPG.  
 XX OS Homo sapiens.  
 XX PN WO9921975-A1.  
 XX PD 06-MAY-1999.  
 XX PF 28-OCT-1998; 98WO-AU000898.  
 XX PR 28-OCT-1997; 97AU-00000062.  
 XX PR 09-DEC-1997; 97AU-00000812.  
 XX PA (AUSU) UNIV AUSTRALIAN NAT.  
 XX PI Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;  
 XX WPI; 1999-312956/26.  
 XX DR N-PSDB; AAX37259.  
 XX KW Polynucleotides encoding mammalian endoglucuronidases, especially  
 PT heparanases, useful to promote wound healing.  
 XX PS Claim 6; Page 69-73; 112pp; English.  
 XX CC The invention relates to nucleic acid sequences that encode heparanase  
 CC enzymes having endoglucuronidase activity. Recombinant heparanases are  
 CC capable of removing the HS side chain from heparan sulfate proteoglycan  
 CC (HSPG). Sulfated oligosaccharides, sulphonates or HSPG can be used to  
 CC inhibit heparanase, this is useful for treatment of a physiological or  
 CC medical condition associated with elevated heparanase activity, such as  
 CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,  
 CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and  
 CC rat heparanases can be used to enhance wound healing, especially  
 CC associated with tissue development and repair. The conditions mentioned  
 CC above can be diagnosed using specific antibodies, and also using primers  
 CC and probes specific for the heparanase polynucleotides. Other uses of the  
 CC heparanases include sequencing sulfated molecules such as HSPG. The  
 CC present sequence represents a human heparanase  
 XX Sequence 543 AA;  
 SQ

Query Match 100.0%; Score 61; DB 2; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 0.084;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTNTDNPYK 10  
 |||||  
 Db 437 CTNTDNPYK 446

RESULT 33  
 AAY57590  
 ID AAY57590 standard; protein; 543 AA.  
 XX AC AAY57590;  
 XX DT 02-MAR-2000 (first entry)  
 XX DE Human heparanase.  
 XX KW Human; heparanase; hpa; genetic modification; expression; anticancer;  
 KW angiogenesis; anti-angiogenic; antiproliferative; antiviral; antitumour;  
 KW anti-atherosclerotic; anti-inflammatory; antineurodegeneration;  
 KW heparan sulphate; heparin-binding growth factor; tumour angiogenesis;  
 KW metastasis; wound healing; restenosis; atherosclerosis; inflammation;  
 KW neurodegeneration; viral infection; cystic fibrosis; cancer; diagnosis;  
 KW micrometastasis; autoimmune lesion; kidney failure.  
 XX OS Homo sapiens.  
 XX PN WO9957244-A1.  
 XX

PD 11-NOV-1999.  
 XX 29-APR-1999; 99WO-US0009256.  
 XX 01-MAY-1998; 98US-00071618.  
 PR 02-MAR-1999; 99US-00260038.  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (FRIE/) FRIEDMAN M M.  
 XX Ben-Artzi H, Ayal-Hershkovitz M, Yacoby-Zeevi O, Pecker I;  
 PI Peleg Y, Shlomi Y;  
 XX WPI; 2000-062144/05.  
 DR N-PSDB; AA239195.  
 XX Engineered cells that express recombinant heparanase, useful  
 PT therapeutically, e.g. for treating angiogenesis and to screen for  
 PT specific inhibitors, potential anticancer agents.  
 XX Claim 3; Page 107-109; 118pp; English.  
 XX The present invention describes genetically modified cells (A) containing  
 CC a polynucleotide (I) that encodes a polypeptide with heparanase activity,  
 CC and express recombinant heparanase (II). Heparanase cleaves heparan  
 CC sulphate (HS) at specific intrachain sites, resulting in release of  
 CC heparin-binding growth factors, enzymes and proteins that are sequestered  
 CC by HS in basement membranes, extracellular matrix or cell surfaces. It  
 CC may also be implicated in tumour angiogenesis and metastases. (II) is  
 CC potentially useful in wound healing and for treating angiogenesis,  
 CC restenosis, atherosclerosis, inflammation, neurodegeneration, viral  
 CC infection and cystic fibrosis. It can also be used to neutralise heparin  
 CC (an alternative to protamine) and to screen for specific inhibitors  
 CC (potentially useful for treating cancer and metastases). Antibodies  
 CC raised against (II) are used for immunodetection and diagnosis of  
 CC micrometastases, autoimmune lesions and kidney failure. (A) provide (II)  
 CC in large quantities, in a form that is homogeneously processed and  
 CC activated/neutralised by a dedicated protease. The present sequence  
 CC represents human heparanase  
 XX  
 SQ Sequence 543 AA;  
 Query Match 100.0%; Score 61; DB 3; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 0.084;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTNTDNPYK 10  
 Db |||||  
 437 CTNTDNPYK 446  
 RESULT 34  
 ID AAB08849  
 AC AAB08849 standard; protein; 543 AA.  
 XX  
 AC AAB08849;  
 XX  
 DT 15-JAN-2001 (first entry)  
 XX Amino acid sequence of a human heparanase polypeptide.  
 DE Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;  
 KW heparin-binding growth factor; cytokine; neurodegenerative plaque;  
 KW wound healing; infection; burn; angiogenesis; restenosis;  
 KW atherosclerosis; inflammation; neurodegenerative disease;  
 KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.  
 XX  
 XX Homo sapiens.  
 OS WO200052178-A1.  
 XX  
 PN 08-SEP-2000.  
 XX  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

PF 14-FEB-2000; 2000WO-US003542.  
 XX 01-MAR-1999; 99US-00258892.  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA (FRIE/) FRIEDMAN M M.  
 XX Pecker I, Vlodavsky I, Feinstein E;  
 PI WPI; 2000-579289/54.  
 DR N-PSDB; AA475051.  
 XX New polynucleotides encoding a polypeptide having heparanase activity,  
 PT useful in wound healing and in gene therapy, particularly in treating  
 PT tumor, inflammation, autoimmunity, neurodegenerative diseases.  
 XX Claim 22; Fig 1; 152pp; English.  
 XX The present sequence represents a human protein with heparanase catalytic  
 CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,  
 CC particularly in treating tumour, inflammation or autoimmunity.  
 CC Particularly, the polynucleotide is useful in modulating the  
 CC bioavailability of heparin-binding growth factors, cellular responses to  
 CC heparin-binding growth factors (e.g. bFGF) and cytokines (e.g.  
 CC interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular  
 CC susceptibility to certain viral and some bacterial and protozoa  
 CC infections, or disintegration of neurodegenerative plaques. The  
 CC polynucleotide is also useful in wound healing (e.g. thermal, chemical or  
 CC radiation burns), and in the treatment of angiogenesis, restenosis,  
 CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-  
 CC Strausler Syndrome or Creutzfeldt-Jakob disease), and some viral,  
 CC bacterial or protozoa infections  
 XX  
 SQ Sequence 543 AA;  
 Query Match 100.0%; Score 61; DB 3; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 0.084;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTNTDNPYK 10  
 Db |||||  
 437 CTNTDNPYK 446  
 RESULT 35  
 AAY52990  
 ID AAY52990 standard; protein; 543 AA.  
 XX  
 AC AAY52990;  
 XX  
 DT 21-FEB-2000 (first entry)  
 XX Human heparanase protein sequence.  
 DE Human; heparanase; hpa; diagnosis; therapy; tumour; cytostatic;  
 KW antidiabetic; immunomodulatory; anti-inflammatory; nephrotropic;  
 KW metastasis; adenocarcinoma; squamous cell carcinoma; teratocarcinoma;  
 KW mesothelioma; melanoma; lymphoma; leukemia; cancer; sepsis; diabetes;  
 KW inflammation; haemorrhagic nephritis; nephrotic syndrome;  
 KW autoimmune disease; anticancer; kidney disease.  
 XX  
 XX Homo sapiens.  
 OS WO9957153-A1.  
 PN 11-NOV-1999.  
 XX  
 PD 29-APR-1999; 99WO-US009255.  
 PF  
 XX 01-MAY-1998; 98US-00071739.  
 PR  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 PA (FRIE/) FRIEDMAN M M.  
 XX  
 PI Pecker I, Vlodavsky I, Friedman Y, Perets T;  
 PI WPI; 2000-052944/04.  
 DR N-PSDB; AA233290.  
 XX Heparanase-specific molecular probes useful for diagnosis and treatment,  
 PT e.g. of tumors, and for targeted drug delivery.  
 XX  
 PS Example; Page 81-82; 90pp; English.  
 XX  
 CC The present invention describes heparanase-specific molecular probes,  
 CC useful for methods of detecting heparanase in situ. The probes and anti-  
 CC heparanase antibodies are used to detect or quantify the expression of  
 CC heparanase, for diagnosis and monitoring of diseases (especially  
 CC metastasis), for treatment of heparanase-associated diseases (e.g.  
 CC tumours, (adeno)carcinoma, squamous cell carcinoma, teratocarcinoma,  
 CC mesothelioma, melanoma, lymphoma or leukemia, a solid cancer for its  
 CC metastases) derived from liver, prostate, bladder, breast, ovary, cervix,  
 CC colon, skin, intestine, stomach, uterus and pancreas, kidney disease,  
 CC diabetes and inflammation, haemorrhagic nephritis, nephrotic syndrome,  
 CC sepsis and inflammatory or autoimmune disease), for targeted drug  
 CC delivery (e.g. of anticancer agents) and as research reagents. The  
 CC present sequence represents human heparanase, which is used in the  
 CC exemplification of the present invention  
 XX  
 XX Sequence 543 AA;  
 SQ

Query Match 100.0%; Score 61; DB 3; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 0.084;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTNTDNPYK 10  
 DB 437 CTNTDNPYK 446  
 |||||

RESULT 36  
 AAY97635  
 ID AAY97635 standard; protein; 543 AA.  
 AC  
 AC AAY97635;  
 XX  
 XX 20-APR-2001 (first entry)  
 DT  
 XX Human heparanase protein sequence.  
 DE  
 XX Heparanase; hnhp1; wound healing; angiogenesis; restenosis; Scrape;  
 KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;  
 KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;  
 KW gene therapy; human.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200100643-A2.  
 PN  
 XX  
 PD 04-JAN-2001.  
 XX  
 XX 19-JUN-2000; 2000WO-IL000358.  
 PF  
 XX 25-JUN-1999; 99US-0140801P.  
 PR  
 XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA  
 XX Pecker I, Michal I, Itzhaki H;  
 PI  
 XX WPI; 2001-137930/14.  
 DR  
 XX New polynucleotides and polypeptides that are distantly homologous to  
 PT heparanase, useful in wound healing, as well as in gene therapy protocols  
 PT for angiogenesis, restenosis, atherosclerosis, or inflammation.

XX Disclosure; Page 64-65; 67pp; English.  
 XX  
 CC This sequence represents a heparanase of the invention. The heparanase  
 CC DNA and protein sequences are useful in wound healing, angiogenesis,  
 CC restenosis, atherosclerosis, inflammation, pulmonary diseases,  
 CC neurodegenerative diseases (such as Scrape, Alzheimer's disease, and  
 CC Creutzfeldt-Jakob disease) or viral infections. The heparanase coding  
 CC sequence is particularly useful in gene therapy  
 XX  
 XX Sequence 543 AA;  
 SQ

Query Match 100.0%; Score 61; DB 4; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 0.084;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTNTDNPYK 10  
 DB 437 CTNTDNPYK 446  
 |||||

RESULT 37  
 AAB86206  
 ID AAB86206 standard; protein; 543 AA.  
 XX  
 AC AAB86206;  
 XX  
 XX 24-AUG-2001 (first entry)  
 DT  
 XX Human heparanase inhibitor protein.  
 DE  
 XX Heparanase; inhibitor; cardiac insufficiency; cardiant; nephrotropic;  
 KW hepatotropic; veterinary medicine; congestive heart failure; dyspnoea;  
 KW primary cardiomyopathy; peripheral odema; pulmonary congestion;  
 KW hepatic congestion; hydrothorax; ascite; nocturia; human.  
 XX  
 OS Homo sapiens.  
 XX  
 XX DE19955803-A1.  
 PN  
 XX 23-MAY-2001.  
 PD  
 XX 19-NOV-1999; 99DE-01055803.  
 PF  
 XX 19-NOV-1999; 99DE-01055803.  
 PR  
 XX (KNOL ) KNOLL AG.  
 PA  
 XX Herr D, Hahn A, Laux V;  
 PI  
 XX WPI; 2001-368371/39.  
 DR  
 XX N-PSDB; AAH20940.  
 XX  
 XX Treatment or prevention of cardiac insufficiency and related conditions,  
 PT e.g. pulmonary congestion and dyspnoea, comprises administration of  
 PT heparanase inhibitor.  
 PT  
 XX Disclosure; Page 11-13; 16pp; German.  
 PS  
 XX This invention describes a novel heparanase inhibitor which can be used  
 CC for the treatment or prevention of cardiac insufficiency and associated  
 CC indications, symptoms and/or malfunctions. The heparanase inhibitor of  
 CC the invention has cardiant, nephrotropic and hepatotropic activity. The  
 CC products of the invention can be used in human and veterinary medicine,  
 CC for the treatment or prevention of congestive heart failure e.g. primary  
 CC cardiomyopathy. Associated conditions treated or prevented with the  
 CC inhibitor are especially peripheral odemas, pulmonary and hepatic  
 CC congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g.  
 CC nocturia can also be treated. This sequence represents the human  
 CC heparanase protein described in the method of the invention  
 XX  
 XX Sequence 543 AA;  
 SQ

Query Match 100.0%; Score 61; DB 4; Length 543;  
Best Local Similarity 100.0%; Pred. No. 0.084;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10  
| | | | | | | | | |  
DB 437 CTNTDNPYK 446

RESULT 38  
AAB88361  
ID AAB88361 standard; protein; 543 AA.  
XX  
AC AAB88361;  
XX  
DT 23-MAY-2001 (first entry)  
XX  
DE Human membrane or secretory protein clone PSEC0090.  
XX  
KW Human; secretory protein; membrane protein; vaccine; gene therapy;  
KW rheumatoid arthritis; diabetes.  
XX  
OS Homo sapiens.  
XX  
PN EP1067182-A2.  
XX  
PD 10-JAN-2001.  
XX  
PF 07-JUL-2000; 2008EP-00114090.  
XX  
PR 08-JUL-1999; 99JP-00194179.  
PR 11-JAN-2000; 2000JP-00118775.  
PR 02-MAY-2000; 2000JP-00183766.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
XX  
DR WPI; 2001-093989/11.  
DR N-PSDB; AAF93788.  
XX  
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in  
PT gene therapy or as candidate target molecules in drug development.  
XX  
PS Claim 1; SEQ ID NO 90; 609pp + Sequence Listing; English.  
XX  
CC This invention relates to nucleic acid sequences AAF93744 - AAF93916  
CC which encode human secretory or membrane proteins represented by AAB88317  
CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and  
CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the  
CC invention. The invention also includes methods for the production of  
CC antibodies directed against the proteins, and cDNA sequences, which can  
CC be used in vaccines. The polynucleotide sequences can be used in gene  
CC therapy. The polynucleotide sequences and the proteins they encode may be  
CC used in the prevention, treatment and diagnosis of diseases associated  
CC with inappropriate secretory protein/membrane protein expression. The  
CC nucleic acids and complementary sequences may also be used as DNA probes  
CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect  
CC and quantitate the presence of similar nucleic acid sequences in samples.  
CC They may also be used to study the expression and function of secretory  
CC proteins/membrane polypeptides and their role in metabolism. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC against them and in assays to identify modulators (agonists and  
CC antagonists) of expression and activity. The antibodies and antagonists  
CC may also be used as therapeutic agents to down regulate expression and  
CC activity. The antibodies may also be used as diagnostic agents for  
CC detecting the presence of the polypeptides in samples (e.g. by enzyme  
CC linked immunosorbent assay (ELISA)). Examples of diseases which may be  
CC treated include rheumatoid arthritis and diabetes  
XX  
SQ Sequence 543 AA;

Query Match 100.0%; Score 61; DB 4; Length 543;

Best Local Similarity 100.0%; Pred. No. 0.084;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10  
| | | | | | | | | |  
DB 437 CTNTDNPYK 446

RESULT 39  
ABB07813  
ID ABB07813 standard; protein; 543 AA.  
XX  
AC ABB07813;  
XX  
DT 03-JUL-2002 (first entry)  
XX  
DE Human heparanase sequence.  
XX  
KW Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;  
KW anti-protozoan; neuroprotective; heparin; human.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..35  
FT /note= "signal peptide"  
FT Protein 36..543  
FT /note= "mature protein"  
XX  
PN US2002034810-A1.  
XX  
PD 21-MAR-2002.  
XX  
PF 16-AUG-2001; 2001US-00930218.  
XX  
PR 20-SEP-2000; 2000US-00666390.  
XX  
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
XX  
PI Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia B;  
XX  
DR WPI; 2002-338926/37.  
XX  
PT Nucleic acid encoding avian and reptile heparanase polypeptide is useful  
PT to treat various heparin-related disorders and the signal peptide is  
PT useful in production of membrane-targeted or secreted recombinant  
PT proteins.  
XX  
PS Disclosure; Fig 1a; 39pp; English.  
XX  
CC The invention relates to an isolated avian and reptile nucleic acid,  
CC encoding a polypeptide with heparanase catalytic activity. The signal  
CC peptide of the nucleic acid can be used to express membrane-associated or  
CC secreted proteins in heterologous expression systems. The encoded  
CC polypeptides can be used to prevent tumour angiogenesis, metastasis and  
CC invasion, and to intervene with pathologies associated with impaired  
CC heparin-binding growth factors, cellular responses to heparin-binding  
CC growth factors and cytokines, cell interaction with plasma lipoproteins,  
CC cellular susceptibility to viral, protozoa and bacterial infections or  
CC disintegration of neurodegenerative plaques. The present sequence  
CC represents a human heparanase protein sequence used in similarity studies  
XX  
SQ Sequence 543 AA;

Query Match 100.0%; Score 61; DB 5; Length 543;  
Best Local Similarity 100.0%; Pred. No. 0.084;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10  
| | | | | | | | | |  
DB 437 CTNTDNPYK 446

RESULT 40  
ADD18950  
ID ADD18950 standard; protein; 543 AA.  
XX AC ADD18950;  
XX DT 15-JAN-2004 (first entry)  
XX DE Human disease related protein SeqID439.  
XX KW human; disease state; cytostatic; antiinflammatory; ophthalmological;  
XX KW antiarteriosclerotic; vulnery; gene therapy;  
XX KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;  
XX KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;  
XX KW glucose transportation; catecholamine synthesis; iron transport;  
XX KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;  
XX KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;  
XX KW inflammatory condition; wound healing.  
XX OS Homo sapiens.  
XX PN W02003018621-A2.  
XX PD 06-MAR-2003.  
XX PF 23-AUG-2002; 2002WO-GB003892.  
XX PR 23-AUG-2001; 2001GB-00020558.  
XX PR 05-OCT-2001; 2001GB-00024037.  
XX PA (OXFO-) OXFORD BIOMEDICA UK LTD.  
XX PI Kingman SM., White J, Ward NR, Harris RA, Naylor S, Mundy CR;  
XX DR WPI; 2003-290046/28.  
XX DR N-PSDB; ADD18951.  
XX PT New substantially purified polypeptide, useful for diagnosing or treating  
XX PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion  
XX PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or  
XX PT wound healing.  
XX PS Claim 25; SEQ ID NO 439; 424pp; English.  
XX CC This invention relates to novel human genes and gene product which are  
XX CC implicated in certain disease states. Compounds which modulate the  
XX CC proteins of the invention may have cytostatic, antiinflammatory, the  
XX CC ophthalmological, antiarteriosclerotic or vulnerary activities. The  
XX CC sequences of the invention may be useful for gene therapy. The invention  
XX CC may be useful for diagnosing or treating a hypoxia-regulated condition,  
XX CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,  
XX CC erythropoiesis, or the biological response to hypoxia conditions  
XX CC including processes such as glycolysis, gluconeogenesis, glucose  
XX CC transportation, catecholamine synthesis, iron transport or nitric oxide  
XX CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion  
XX CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,  
XX CC inflammatory conditions or wound healing. The present sequence is that of  
XX CC a disease related protein of the invention.

XX AC ADG88800;  
XX DT 11-MAR-2004 (first entry)  
XX DE Human hpa protein.  
XX KW Wound healing; heparanase; ulcer; burn; laceration; surgical incision;  
XX KW necrosis; pressure wound; diabetic ulcer; angiogenesis; human; therapy.  
XX OS Homo sapiens.  
XX PN US2003161823-A1.  
XX PD 28-AUG-2003.  
XX PF 14-JAN-2003; 2003US-00341582.  
XX PR 31-AUG-1998; 98WO-US017954.  
XX PR 01-MAR-1999; 99US-00258892.  
XX PR 06-FEB-2001; 2001US-00776874.  
XX PR 05-SEP-2001; 2001WO-IL000830.  
XX PR 19-NOV-2001; 2001US-00988113.  
XX PA (ILAN/) ILAN N.  
XX PA (VL0D/) VL0DAVSKY I.  
XX PA (YACO/) YACOBY-ZEEVI O.  
XX PA (PECK/) PECKER I.  
XX PA (FEIN/) FEINSTEIN E.  
XX PI Ilan N, Vlodavsky I, Yacoby-Zeevi O, Pecker I, Feinstein E;  
XX DR WPI; 2003-897910/82.  
XX DR N-PSDB; ADG88799, ADG88801, ADG88832.  
XX PT Composition for treating a wound comprising recombinant heparanase is  
XX PT useful to induce or accelerate wound healing and induce or accelerate  
XX PT angiogenesis.  
XX PS Claim 2; SEQ ID NO 10; 143pp; English.  
XX CC The present invention relates to methods and compositions for inducing  
XX CC and/or accelerating wound healing via the catalytic activity of  
XX CC heparanase. The invention is used to induce or accelerate a healing  
XX CC process, particularly of an ulcer, burn, laceration, surgical incision,  
XX CC necrosis, pressure wound, diabetic ulcer and to induce or accelerate  
XX CC angiogenesis. The present sequence is human hpa protein.  
XX SQ Sequence 543 AA;  
Query Match 100.0%; Score 61; DB 7; Length 543;  
Best Local Similarity 100.0%; Pred. No. 0.084;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CTNTDNPYK 10  
Db 437 CTNTDNPYK 446  
RESULT 42  
ADL16379  
ID ADL16379 standard; protein; 543 AA.  
XX AC ADL16379;  
XX DT 06-MAY-2004 (first entry)  
XX DE Human heparanase partial protein.  
XX KW heparanase; heparanase-dependent cancer; cancer;  
XX KW autoimmune reaction; inflammation; chromosome 4; enzyme.  
XX OS Homo sapiens.



XX US2003236215-A1.  
 XX 25-DEC-2003.  
 XX 09-JUN-2003; 2003US-00456573.  
 XX 31-AUG-1998; 98WO-US017954.  
 XX 01-MAR-1999; 99US-00258892.  
 XX 08-NOV-1999; 99US-00435739.  
 XX (INST-) INSTIGHT STRATEGY & MARKETING LTD.  
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX Pecker I, Vlodavsky I, Feinstein E;  
 XX WPI; 2004-070610/07.  
 XX New antisense oligonucleotide hybridizable with a polynucleotide encoding  
 PT a polypeptide with heparanase activity, useful for treating diseases such  
 PT as cancer and autoimmune disorders.  
 XX Claim 3; SEQ ID NO 10; 108pp; English.  
 XX The invention relates to an antisense oligonucleotide (ASO) comprising a  
 CC polynucleotide or a polynucleotide analogue of at least 10 bases being  
 CC hybridisable in vivo, under physiological conditions, with a portion of  
 CC a polynucleotide strand encoding a polypeptide having heparanase  
 CC catalytic activity. Also included are a method of in vivo downregulating  
 CC heparanase activity (comprising administering the ASO in vivo), a method  
 CC of treating a subject suffering from a pathological condition  
 CC (characterised by heparanase activity, comprising administering ASO to  
 CC the subject), a pharmaceutical composition comprising the ASO and a  
 CC carrier, an antisense nucleic acid construct (comprising a promoter  
 CC sequence and a polynucleotide sequence directing the synthesis of an  
 CC antisense RNA sequence of at least 10 bases being hybridisable in vivo,  
 CC under physiological conditions, with a polynucleotide strand encoding a  
 CC polypeptide having heparanase catalytic activity), a method of in vivo  
 CC downregulating heparanase activity (comprising administering in vivo the  
 CC antisense nucleic acid construct), a pharmaceutical composition  
 CC comprising the antisense nucleic acid construct and a carrier, and an  
 CC antisense oligonucleotide comprising a polynucleotide or a polynucleotide  
 CC analogue of at least 10 bases being hybridisable in vivo, under  
 CC physiological conditions, with a portion of a polynucleotide strand being  
 CC characterised by forming at least a portion of an untranslated region  
 CC (UTR) for a polynucleotide strand encoding a polypeptide having  
 CC heparanase catalytic activity. The methods and compositions of the  
 CC present invention are useful for the prevention and/or treatment of  
 CC diseases or conditions associated with aberrant heparanase activity, such  
 CC as heparanase-dependent cancer, cancer, autoimmune reaction and  
 CC inflammation. The gene for human heparanase is located on chromosome 4.  
 CC The present sequence is a human heparanase protein.  
 XX SQ Sequence 543 AA;  
 Query Match 100.0%; Score 61; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 0.084;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTNTDNPYK 10  
 |||||  
 Db 437 CTNTDNPYK 446

RESULT 43  
 ADK52086  
 ID ADK52086 standard; protein; 543 AA.  
 XX AC ADK52086;  
 XX 20-MAY-2004 (first entry)  
 DT Human atopic dermatitis/psoriasis-associated protein #1.  
 DE

XX Human; atopic dermatitis; psoriasis; dermatological; anti-inflammatory;  
 KW antipsoriatic; rash.  
 XX OS Homo sapiens.  
 XX WO2004016785-A1.  
 XX 26-FEB-2004.  
 XX 06-AUG-2003; 2003WO-JP009999.  
 XX 06-AUG-2002; 2002JP-00229319.  
 XX 14-MAY-2003; 2003JP-00136544.  
 XX (GENO-) GENOX RES INC.  
 XX (UJJU-) UNIV JUNTENDO.  
 XX Itoh M, Ogawa K, Shinagawa A, Sudo H, Ogawa H, Ra C;  
 PI Mitsuishi K;  
 XX WPI; 2004-214514/20.  
 XX N-PSDB; ADK51968.  
 XX Detecting atopic dermatitis or psoriasis comprises assaying levels of  
 PT expression of an indicator gene at a rash site and non-rash site of a  
 PT person with atopic dermatitis or psoriasis.  
 XX Example 2; SEQ ID NO 119; 484pp; Japanese.  
 XX The invention relates to detecting atopic dermatitis or psoriasis  
 CC comprising assaying the levels of expression of an indicator gene at a  
 CC rash site and non-rash site of a person with atopic dermatitis or  
 CC psoriasis, comparing these levels with those of a healthy person, and  
 CC determining that if the levels of indicators are higher or lower, then  
 CC this indicates the disease. Also included are a reagent for detecting  
 CC atopic dermatitis or psoriasis, a kit for screening for treatments, a  
 CC transgenic non human vertebrate animal models for the diseases, an agent  
 CC for inducing the diseases in mice and a DNA chip for assaying for the  
 CC indicator genes. The method is used for treatment, detection and animal  
 CC models for research of atopic dermatitis and psoriasis. The present  
 CC sequence is a protein encoded by an indicator gene of the invention.  
 XX SQ Sequence 543 AA;  
 Query Match 100.0%; Score 61; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 0.084;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTNTDNPYK 10  
 |||||  
 Db 437 CTNTDNPYK 446

RESULT 44  
 ADM48716  
 ID ADM48716 standard; protein; 543 AA.  
 XX AC ADM48716;  
 XX 03-JUN-2004 (first entry)  
 DT Human hpa protein #1.  
 DE Transgenic animal; heparanase; cancer; viral infection; restenosis;  
 KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;  
 KW human.  
 XX OS Homo sapiens.  
 XX US2003217375-A1.  
 XX 20-NOV-2003.

```

XX PF 24-FEB-2003; 2003US-00371218.
XX PR 31-AUG-1998; 98WO-US017954.
XX PR 01-MAR-1999; 99US-00258892.
XX PR 06-FEB-2001; 2001US-00776874.
XX PR 19-NOV-2001; 2001US-00988113.
XX PA (ZCHA/) ZCHARIA E.
XX PA (VLOD/) VLODAVSKY I.
XX PA (METZ/) METZGER S.
XX PA (PECK/) PECKER I.
XX PA (ILAN/) ILAN N.
XX PA (CHAJ/) CHAJEK-SHAUL T.
XX PA (GOLD/) GOLDSCHMIDT O.
XX Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;
XX Chajek-Shaul T, Goldshmidt O;
XX WPI; 2004-021918/02.
XX DR N-PSDB; ADM48715, ADM48717.
XX PT New transgenic non-human animal expressing heparinase, useful as models
XX PT for human disease, such as cancers, viral infection, neurodegenerative
XX PT diseases, restenosis, atherosclerosis and pulmonary disorders.
XX PS Example 1; SEQ ID NO 10; 106pp; English.
XX CC The present invention relates to a transgenic non-human animal whose
XX CC genome comprises an exogenous polynucleotide sequence, including a
XX CC promoter active in tissues of the non-human, a region encoding a human
XX CC heparanase, where the promoter and the region encoding human heparanase
XX CC are operably linked in the exogenous polynucleotide such that human
XX CC heparanase is expressed in at least a portion of the cells of the non-
XX CC human animal. The methods and compositions of the present invention are
XX CC useful for the production of transgenic animals expressing heparanase, to
XX CC be used as models for human diseases such as cancers, viral infection,
XX CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary
XX CC disorders. The present sequence is human hpa protein used in the
XX CC exemplification of the invention.
XX SQ Sequence 543 AA;
XX Query Match 100.0%; Score 61; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 0.084;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 CTNTDNPYK 10
XX DB |||||
XX 437 CTNTDNPYK 446
XX RESULT 45
XX ADM48759
XX ID ADM48759 standard; protein; 543 AA.
XX AC ADM48759;
XX XX
XX DT 03-JUN-2004 (first entry)
XX DE Human hpa protein #2.
XX KW Transgenic animal; heparanase; cancer; viral infection; restenosis;
XX KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;
XX KW human.
XX OS Homo sapiens.
XX PN US2003217375-A1.
XX PD 20-NOV-2003.
XX PF 24-FEB-2003; 2003US-00371218.

```

```

XX PR 31-AUG-1998; 98WO-US017954.
XX PR 01-MAR-1999; 99US-00258892.
XX PR 06-FEB-2001; 2001US-00776874.
XX PR 19-NOV-2001; 2001US-00988113.
XX PA (ZCHA/) ZCHARIA E.
XX PA (VLOD/) VLODAVSKY I.
XX PA (METZ/) METZGER S.
XX PA (PECK/) PECKER I.
XX PA (ILAN/) ILAN N.
XX PA (CHAJ/) CHAJEK-SHAUL T.
XX PA (GOLD/) GOLDSCHMIDT O.
XX Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;
XX Chajek-Shaul T, Goldshmidt O;
XX WPI; 2004-021918/02.
XX DR N-PSDB; ADM48748.
XX PT New transgenic non-human animal expressing heparinase, useful as models
XX PT for human disease, such as cancers, viral infection, neurodegenerative
XX PT diseases, restenosis, atherosclerosis and pulmonary disorders.
XX PS Example 10; Fig 16; 106pp; English.
XX CC The present invention relates to a transgenic non-human animal whose
XX CC genome comprises an exogenous polynucleotide sequence, including a
XX CC promoter active in tissues of the non-human, a region encoding a human
XX CC heparanase, where the promoter and the region encoding human heparanase
XX CC are operably linked in the exogenous polynucleotide such that human
XX CC heparanase is expressed in at least a portion of the cells of the non-
XX CC human animal. The methods and compositions of the present invention are
XX CC useful for the production of transgenic animals expressing heparanase, to
XX CC be used as models for human diseases such as cancers, viral infection,
XX CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary
XX CC disorders. The present sequence is human hpa protein used in the
XX CC exemplification of the invention.
XX SQ Sequence 543 AA;
XX Query Match 100.0%; Score 61; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 0.084;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 CTNTDNPYK 10
XX DB |||||
XX 437 CTNTDNPYK 446
XX RESULT 46
XX ADN05074
XX ID ADN05074 standard; protein; 543 AA.
XX AC ADN05074;
XX XX
XX DT 01-JUL-2004 (first entry)
XX DE Antipsoriatic protein sequence #716.
XX KW antipsoriatic; gene therapy; psoriasis; diagnosis.
XX OS Homo sapiens.
XX PN WO2004028479-A2.
XX PD 08-APR-2004.
XX PF 25-SEP-2003; 2003WO-US030907.
XX PR 25-SEP-2002; 2002US-0414006P.
XX PA (GETH ) GENENTECH INC.

```

XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;  
PI Wu TD;  
XX WPI; 2004-305105/28.  
XX N-PSDB; ADN05073.  
XX New PRO nucleic acid or polypeptide, useful for preparing a  
PT pharmaceutical composition for diagnosing or treating psoriasis in a  
PT mammal.  
XX Claim 9; SEQ ID NO 1468; 3069pp; English.  
XX The invention relates to novel polynucleotide and polypeptides for  
CC treating psoriasis or a sequence having at least 80% identity to the  
CC above sequences. The nucleic acid is useful for preparing a composition  
CC for diagnosing or treating psoriasis in a mammal. This sequence  
CC corresponds to one of the polypeptides of the invention.  
XX SQ Sequence 543 AA;  
Query Match 100.0%; Score 61; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 0.084;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CTNTDNPYK 10  
DB 437 CTNTDNPYK 446  
RESULT 47  
ID ADN04902 standard; protein; 543 AA.  
AC ADN04902;  
XX 01-JUL-2004 (first entry)  
DT Antipsoriatic protein sequence #631.  
DE antipsoriatic; gene therapy; psoriasis; diagnosis.  
XX Homo sapiens.  
XX WO2004028479-A2.  
XX 08-APR-2004.  
XX 25-SEP-2003; 2003WO-US030907.  
XX 25-SEP-2002; 2002US-0414006P.  
XX (GETH ) GENENTECH INC.  
XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;  
PI Wu TD;  
XX WPI; 2004-305105/28.  
XX N-PSDB; ADN04901.  
XX New PRO nucleic acid or polypeptide, useful for preparing a  
PT pharmaceutical composition for diagnosing or treating psoriasis in a  
PT mammal.  
XX Claim 9; SEQ ID NO 1296; 3069pp; English.  
XX The invention relates to novel polynucleotide and polypeptides for  
CC treating psoriasis or a sequence having at least 80% identity to the  
CC above sequences. The nucleic acid is useful for preparing a composition  
CC for diagnosing or treating psoriasis in a mammal. This sequence  
CC corresponds to one of the polypeptides of the invention.  
XX SQ Sequence 543 AA;

Query Match 100.0%; Score 61; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 0.084;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CTNTDNPYK 10  
DB 437 CTNTDNPYK 446  
RESULT 48  
ID ADO63831 standard; protein; 543 AA.  
AC ADO63831;  
XX 26-AUG-2004 (first entry)  
DT Human heparanase mutant E378A.  
DE Human; heparanase; heparanase-derived protein; heparanase mutant;  
XX non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;  
KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;  
KW vascular graft; platelet adhesion; platelet aggregation;  
KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
KW vulnary; mutant; mutein; enzyme.  
XX Homo sapiens.  
XX Synthetic.  
FH Key Location/Qualifiers  
FT Active-site /note= "Aactive site proton donor"  
FT Active-site /note= "Active site nucleophile"  
FT Misc-difference 378 /note= "Ala replaces wild-type Glu"  
XX WO2004048558-A2.  
XX 10-JUN-2004.  
XX 24-NOV-2003; 2003WO-IL000989.  
XX 24-NOV-2002; 2002IL-00153059.  
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
XX Vlodayvsky I, Zecharia E, Goldshmidt O, Ilan N;  
XX WPI; 2004-450373/42.  
XX New nucleic acid construct comprising heparanase-derived polypeptide,  
PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
PT thrombasthenia, or Bernard-Soulier syndrome.  
XX Example 4; Page; 128pp; English.  
XX The invention relates to nucleic acid constructs comprising a nucleic  
CC acid encoding a heparanase-derived protein which lacks heparanase  
CC endoglycosidase catalytic activity but which retains its cell-cell and  
CC cell-matrix adhesion properties. The constructs of the invention  
CC optionally further comprise operably linked regulatory elements. The  
CC invention also relates to the heparanase-derived proteins and host cells  
CC comprising the nucleic acid constructs of the invention. The heparanase-  
CC derived proteins are especially mutants of human heparanase in which the  
CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
CC host cells are useful in preparing a tissue sealant composition for  
CC sealing injuries, reducing the loss of blood, accelerating the healing

CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialisation of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be used in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte  
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be used to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents a human  
 CC heparanase mutant E378A created in an example of the invention which  
 CC retains its heparanase catalytic activity. The present sequence is not  
 CC shown in the invention, but is derived from the protein sequence of  
 CC Genbank accession number AF14325 and the information provided on page  
 CC 70.

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 61; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 0.084;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10  
 |||||  
 Db 437 CTNTDNPYK 446

RESULT 49  
 ADO63824

ID ADO63824 standard; protein; 543 AA.  
 XX ADO63824;  
 XX 26-AUG-2004 (first entry)  
 XX Human heparanase mutant E225A/E343A, SEQ ID:9.

KW Human; heparanase; heparanase-derived protein; heparanase mutant;  
 KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;  
 KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;  
 KW vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnery; mutant; mutein.

OS Homo sapiens.  
 OS Synthetic.

XX Key Location/Qualifiers  
 FT Misc-difference 225 /note= "Ala replaces wild-type Glu (active site proton  
 FT donor)"  
 FT Misc-difference 343 /note= "Ala replaces wild-type Glu (active site  
 FT nucleophile)"  
 XX WO2004048558-A2.  
 XX 10-JUN-2004.  
 XX 24-NOV-2003; 2003WO-IL000989.  
 XX 24-NOV-2002; 2002IL-00153059.  
 XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
 XX WPI; 2004-450373/42.  
 XX N-PSDB; ADO63818.  
 XX New nucleic acid construct comprising heparanase-derived polypeptide,

PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
 XX thrombasthenia, or Bernard-Soulier syndrome.  
 PS Claim 9; SEQ ID NO 9; 128pp; English.

XX The invention relates to nucleic acid constructs comprising a nucleic  
 CC acid encoding a heparanase-derived protein which lacks heparanase  
 CC endoglycosidase catalytic activity but which retains its cell-cell and  
 CC cell-matrix adhesion properties. The constructs of the invention  
 CC optionally further comprise operably linked regulatory elements. The  
 CC invention also relates to the heparanase-derived proteins and host cells  
 CC comprising the nucleic acid constructs of the invention. The heparanase-  
 CC derived proteins are especially mutants of human heparanase in which the  
 CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
 CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
 CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
 CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
 CC host cells are useful in preparing a tissue sealant composition for  
 CC sealing injuries, reducing the loss of blood, accelerating the healing  
 CC and homeostasis of an injury, accelerating blood vessel endothelium  
 CC formation or the endothelialisation of vascular grafts, accelerating the  
 CC adhesive activity of mammalian cells, and accelerating the adhesion and  
 CC aggregation of platelets. They may also be used in the treatment of  
 CC disorders associated with adhesion deficiency such as LAD (leukocyte  
 CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
 CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
 CC cells of the invention may additionally be used to screen for modulators of  
 CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
 CC synthetic vascular graft comprising a tube made of a biocompatible  
 CC material lined with the cells. The present sequence represents the human  
 CC heparanase double mutant E225A/E343A.

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 61; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 0.084;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10  
 |||||  
 Db 437 CTNTDNPYK 446

RESULT 50  
 ADO63823

ID ADO63823 standard; protein; 543 AA.  
 XX ADO63823;  
 XX 26-AUG-2004 (first entry)  
 XX Human heparanase mutant E343A, SEQ ID:8.

KW Human; heparanase; heparanase-derived protein; heparanase mutant;  
 KW non-catalytic; enzymatically inactive; cell adhesion; matrix adhesion;  
 KW tissue sealant; injury; blood loss; endothelialisation; blood vessel;  
 KW vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnery; mutant; mutein.

OS Homo sapiens.  
 OS Synthetic.

XX Key Location/Qualifiers  
 FT Active-site 225 /note= "Active site proton donor"  
 FT Misc-difference 343 /note= "Ala replaces wild-type Glu (active site  
 FT nucleophile)"  
 XX WO2004048558-A2.

PD 10-JUN-2004.  
XX  
XX  
XX 24-NOV-2003; 2003WO-IL000989.  
XX  
XX  
XX 24-NOV-2002; 2002IL-00153059.  
XX  
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
XX  
XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
XX  
XX WPI; 2004-450373/42.  
XX  
XX N-PSDB; ADO63817.  
XX  
XX New nucleic acid construct comprising heparanase-derived polypeptide,  
PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
PT thrombasthenia, or Bernard-Soulier syndrome.  
XX  
XX  
XX Claim 9; SEQ ID NO 8; 128pp; English.  
XX  
XX The invention relates to nucleic acid constructs comprising a nucleic  
CC acid encoding a heparanase-derived protein which lacks heparanase  
CC endoglycosidase catalytic activity but which retains its cell-cell and  
CC cell-matrix adhesion properties. The constructs of the invention  
CC optionally further comprise operably linked regulatory elements. The  
CC invention also relates to the heparanase-derived proteins and host cells  
CC comprising the nucleic acid constructs of the invention. The heparanase-  
CC derived proteins are especially mutants of human heparanase in which the  
CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
CC host cells are useful in preparing a tissue sealant composition for  
CC sealing injuries, reducing the loss of blood, accelerating the healing  
CC and homeostasis of an injury, accelerating blood vessel endothelium  
CC formation or the endothelialisation of vascular grafts, accelerating the  
CC adhesive activity of mammalian cells, and accelerating the adhesion and  
CC aggregation of platelets. They may also be use in the treatment of  
CC disorders associated with adhesion deficiency such as LAD (leukocyte  
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
CC cells of the invention may additionally be to screen for modulators of  
CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
CC synthetic vascular graft comprising a tube made of a biocompatible  
CC material lined with the cells. The present sequence represents the human  
CC heparanase mutant E343A.  
XX  
XX Sequence 543 AA;  
SQ  
Query Match 100.0%; Score 61; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 0.084;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CTNTDNPYK 10  
| | | | | | | | | |  
Db 437 CTNTDNPYK 446  
Search completed: June 5, 2006, 12:42:54  
Job time : 96.589 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:43:17 ; Search time 11.0959 Seconds  
(without alignments)  
86.714 Million cell updates/sec

Title: US-10-645-659A-6

Perfect score: 61

Sequence: 1 CTNTDNPYK 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : PIR 80.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	70.5	868	2 A83547	probable aconitate
2	43	70.5	868	2 B82213	aconitate hydratase
3	43	70.5	868	2 C81200	aconitate hydratase
4	43	70.5	868	2 H81775	aconitate hydratase
5	41	67.2	87	2 S01982	hypothetical prote
6	40	65.6	297	2 F97218	dihydrodipicolinat
7	39	63.9	259	2 F97218	glycosyltransferas
8	38	62.3	384	2 T44652	UDP-N-acetylglucos
9	38	62.3	487	2 T49424	hypothetical prote
10	38	62.3	616	2 T07611	aconitate hydratase
11	38	62.3	848	2 G86708	aconitate hydratase
12	38	62.3	855	2 C90262	aconitate hydratase
13	38	62.3	870	2 E72541	probable aconitate
14	38	62.3	878	2 A97854	aconitate hydratase
15	38	62.3	878	2 A71641	aconitate hydratase
16	38	62.3	887	2 T27868	hypothetical prote
17	38	62.3	889	1 A44153	aconitate hydratase
18	38	62.3	889	2 A44154	aconitate hydratase
19	38	62.3	889	2 S26403	aconitate hydratase
20	38	62.3	889	2 S18720	aconitate hydratase
21	38	62.3	890	2 AH0270	aconitate hydratase
22	38	62.3	891	2 AH0654	aconitate hydratase
23	38	62.3	891	2 B48642	aconitate hydratase
24	38	62.3	891	2 A90860	aconitate hydratase
25	38	62.3	891	2 E85759	aconitate hydratase
26	38	62.3	891	2 G84875	aconitate hydratase
27	38	62.3	895	2 A13483	aconitate hydratase
28	38	62.3	897	2 AD2906	aconitate hydratase
29	38	62.3	897	2 E97681	aconitate hydratase

30	38	62.3	898	2 B84471	cytoplasmic aconit
31	38	62.3	898	2 T04693	aconitate hydratase
32	38	62.3	898	2 T10101	aconitate hydratase
33	38	62.3	900	2 A11279	aconitate hydratase
34	38	62.3	900	2 A11642	aconitate hydratase
35	38	62.3	901	2 F89910	aconitate hydratase
36	38	62.3	903	2 A87704	aconitate hydratase
37	38	62.3	906	2 G75362	aconitate hydratase
38	38	62.3	907	2 T04820	aconitate hydratase
39	38	62.3	908	2 G82824	aconitate hydratase
40	38	62.3	909	2 G69599	aconitate hydratase
41	38	62.3	910	2 B83451	aconitate hydratase
42	38	62.3	943	2 F70873	aconitate hydratase
43	38	62.3	944	2 G87135	aconitate hydratase
44	37	60.7	201	2 A12205	hypothetical prote
45	37	60.7	323	2 A96670	hypothetical prote
46	37	60.7	661	2 I52603	MPS1 protein - mou
47	37	60.7	756	2 F64112	malate dehydrogena
48	37	60.7	805	2 E88412	protein C44P1.5 li
49	37	60.7	805	2 T19936	hypothetical prote
50	37	60.7	1260	2 T37523	probable oxoprolin
51	37	60.7	1470	2 S45323	genome polyprotein
52	36	59.0	151	2 P00506	hypothetical prote
53	36	59.0	604	2 D90523	lipoprotein (impor
54	36	59.0	728	2 H65208	hypothetical 82.6
55	36	59.0	728	2 E86093	regulator of acety
56	36	59.0	728	2 G91245	regulator of acety
57	36	59.0	865	2 T41685	probable gamma-ada
58	36	59.0	1028	2 A53449	plasmacytoma-assoc
59	36	59.0	1028	2 I58164	BIG-1 protein - ra
60	35	57.4	117	2 F72536	hypothetical prote
61	35	57.4	343	2 I69009	MHC class I RT1.C-
62	35	57.4	384	2 T19265	hypothetical prote
63	35	57.4	402	2 T37879	probable glutamate
64	35	57.4	415	2 S39195	levansucrase - Erw
65	35	57.4	423	2 S46741	glutamate 5-kinase
66	35	57.4	428	2 S61186	glutamate 5-kinase
67	35	57.4	610	1 I46001	C4b-binding protei
68	35	57.4	739	2 F86337	F14010.2 protein -
69	35	57.4	813	2 T02672	hypothetical prote
70	35	57.4	868	2 G64119	DNA topoisomerase
71	35	57.4	875	2 B97236	probable permease
72	35	57.4	914	1 JN0550	iodide peroxidase
73	35	57.4	914	1 S07047	iodide peroxidase
74	35	57.4	915	2 A49874	metabotropic gluta
75	35	57.4	926	1 OPPGIT	iodide peroxidase
76	35	57.4	933	1 OPHUIT	iodide peroxidase
77	35	57.4	1215	2 E70614	hypothetical prote
78	34	55.7	72	2 G97134	hypothetical prote
79	34	55.7	98	2 H34964	Ig heavy chain V-I
80	34	55.7	98	2 A49051	Ig heavy chain V-I
81	34	55.7	117	2 S18554	Ig heavy chain V r
82	34	55.7	120	2 S26789	Ig heavy chain V r
83	34	55.7	131	2 S26792	Ig heavy chain V r
84	34	55.7	136	2 T48077	hypothetical prote
85	34	55.7	275	1 B64077	bis(5'-nucleosyl)-
86	34	55.7	289	2 B66649	hypothetical prote
87	34	55.7	321	2 H69869	transcription regu
88	34	55.7	341	2 T16951	hypothetical prote
89	34	55.7	374	2 S07845	mRNA maturase b13
90	34	55.7	429	2 T16656	hypothetical prote
91	34	55.7	469	2 S44620	C50C3.1 protein -
92	34	55.7	485	2 C71400	adenosylhomocyste
93	34	55.7	517	2 S21209	mRNA maturase b13
94	34	55.7	517	2 S78664	mRNA maturase b13
95	34	55.7	599	1 A54906	afamin precursor -
96	34	55.7	682	2 T12968	hypothetical prote
97	34	55.7	770	2 T23999	hypothetical prote
98	34	55.7	1074	2 T24877	hypothetical prote
99	34	55.7	1076	2 T24887	hypothetical prote
100	34	55.7	1402	2 T24664	hypothetical prote
101	34	55.7	1437	2 S07430	M polyprotein prec
102	34	55.7	2135	2 T14602	variant-specific s

103	33.5	54.9	680	2	D64420	N-methylhydantoina	176	32	52.5	173	2	AD0718	probable lipoprote
104	33.5	54.9	1002	2	T30546	major surface glyc	177	32	52.5	179	2	BH3830	peptide methionine
105	33.5	54.9	1162	2	B97852	hypothetical prote	178	32	52.5	184	2	B90592	hypothetical prote
106	33.5	54.9	1169	2	T31639	hypothetical prote	179	32	52.5	205	2	T15757	hypothetical prote
107	33	54.1	103	2	F83918	hypothetical prote	180	32	52.5	229	2	S44217	coat protein - tom
108	33	54.1	103	2	T21100	hypothetical prote	181	32	52.5	241	1	QOCVW2	coat protein - abu
109	33	54.1	164	2	G71427	hypothetical prote	182	32	52.5	241	2	G90190	conserved hypotet
110	33	54.1	179	2	PN0096	hypothetical 19K p	183	32	52.5	244	2	JC8019	CD58 protein - pig
111	33	54.1	181	2	A31245	outer spore coat p	184	32	52.5	247	1	QOCVR1	coat protein - tom
112	33	54.1	185	2	G64088	transcription anti	185	32	52.5	251	1	JQ1869	coat protein - tom
113	33	54.1	187	1	G70406	siroheme synthase	186	32	52.5	251	1	QOCVP2	coat protein - pot
114	33	54.1	224	1	A46343	nonstructural prot	187	32	52.5	251	1	QOCVS2	coat protein - squ
115	33	54.1	224	1	JQ0597	nonstructural prot	188	32	52.5	251	2	QOCVS2	coat protein - pep
116	33	54.1	225	2	A31926	transcription acti	189	32	52.5	251	2	S31874	ARI protein - pepp
117	33	54.1	245	2	S48363	hypothetical prote	190	32	52.5	256	1	JQ1886	coat protein - tom
118	33	54.1	252	2	S08054	hypothetical prote	191	32	52.5	256	2	F97163	flagellar basal bo
119	33	54.1	277	2	D89455	protein F55A4.2 [i	192	32	52.5	256	2	JQ2326	coat protein - Ind
120	33	54.1	290	2	T37151	probable DNA-bindi	193	32	52.5	256	2	S22589	hypothetical prote
121	33	54.1	330	2	I69645	probable nucleotid	194	32	52.5	256	2	S58346	coat protein VI -
122	33	54.1	334	2	F82038	adenosine deaminas	195	32	52.5	257	2	JQ2332	ARI protein - cass
123	33	54.1	350	2	T39555	hypothetical prote	196	32	52.5	257	2	S39210	gene V2 protein -
124	33	54.1	372	2	A83167	conserved hypotet	197	32	52.5	257	2	S39234	gene V2 protein -
125	33	54.1	393	1	A48573	calreticulin autoa	198	32	52.5	258	1	QOCMCI	coat protein - cas
126	33	54.1	395	2	D85570	hypothetical prote	199	32	52.5	258	1	VCOMCN	coat protein - cas
127	33	54.1	428	2	G81504	enolase CP1071 [im	200	32	52.5	258	2	S25624	coat protein - cas
128	33	54.1	428	2	A72034	enolase - Chlamydo	201	32	52.5	258	2	S25622	coat protein - cas
129	33	54.1	428	2	F86590	enolase [imported]	202	32	52.5	258	2	S25623	coat protein - cas
130	33	54.1	458	2	A12722	tail fiber protein	203	32	52.5	260	1	QOCVCL	coat protein - tom
131	33	54.1	459	2	T08713	NAD+ ADP-ribosyltr	204	32	52.5	263	2	T16329	hypothetical prote
132	33	54.1	462	2	E97504	gene 17, tail fibre	205	32	52.5	280	2	D82097	methionine aminope
133	33	54.1	485	2	T06764	adenosylhomocyste	206	32	52.5	285	2	AC1537	hypothetical prote
134	33	54.1	504	2	I57022	integral membrane	207	32	52.5	316	2	E44636	NAM (no apical mer
135	33	54.1	521	2	S02018	regulatory protein	208	32	52.5	325	2	F82350	lipid A biosynthes
136	33	54.1	535	2	C95040	glucan 1,6-alpha-9	209	32	52.5	325	2	A71164	hypothetical prote
137	33	54.1	537	2	F90465	medium-chain-fatty	210	32	52.5	332	2	H82340	gluconate utilizac
138	33	54.1	583	2	T20291	hypothetical prote	211	32	52.5	335	2	T31730	hypothetical prote
139	33	54.1	604	2	H89914	hypothetical prote	212	32	52.5	335	2	T32657	hypothetical prote
140	33	54.1	611	2	A54086	calnexin-t - mouse	213	32	52.5	341	2	D75068	glutamine-fructose
141	33	54.1	611	2	A53418	calnexin precursor	214	32	52.5	388	2	S48879	protein kinase SMK
142	33	54.1	629	2	T39285	probable transmemb	215	32	52.5	393	2	S06259	major outer membra
143	33	54.1	641	2	T52489	Rec10 protein [val	216	32	52.5	394	2	S11012	major outer membra
144	33	54.1	652	2	S25265	outer membrane pro	217	32	52.5	395	2	F86740	teichoic acid bios
145	33	54.1	652	2	D82317	iron-regulated out	218	32	52.5	399	2	B83398	hypothetical prote
146	33	54.1	692	2	A97013	hypothetical prote	219	32	52.5	408	2	T16601	hypothetical prote
147	33	54.1	718	2	A51832	ATP-dependent DNA	220	32	52.5	428	2	S09134	gene ND4L intron 1
148	33	54.1	770	2	S60676	cellobiose oxidase	221	32	52.5	429	2	T15303	hypothetical prote
149	33	54.1	791	2	S55725	rec10 protein - fi	222	32	52.5	439	2	T30787	hypothetical prote
150	33	54.1	803	2	F90485	hypothetical prote	223	32	52.5	439	2	T28472	hypothetical prote
151	33	54.1	837	2	AB3383	hypothetical membr	224	32	52.5	439	2	H72154	SIOL protein - var
152	33	54.1	872	2	H86435	protein F17F8.5 [i	225	32	52.5	439	2	F42507	F10L protein - vac
153	33	54.1	889	2	C72565	probable valyl-tRN	226	32	52.5	439	2	E36840	C14L protein - var
154	33	54.1	882	2	E88465	protein B0244.6 [i	227	32	52.5	446	2	T00846	hypothetical prote
155	33	54.1	983	2	E86989	probable integral	228	32	52.5	470	2	S76942	hypothetical prote
156	33	54.1	991	2	T48631	polynucleotide pho	229	32	52.5	472	2	A12997	transcription regu
157	33	54.1	1037	2	E81980	pilus-associated p	230	32	52.5	482	2	H98285	probable transcrip
158	33	54.1	1050	2	JC7809	heparinase (EC 3.2	231	32	52.5	492	2	T21934	hypothetical prote
159	33	54.1	1060	2	T43046	retrovirus-related	232	32	52.5	510	2	E82374	threonine ammonia-
160	33	54.1	1150	2	T13824	LK6 protein kinase	233	32	52.5	514	2	T04085	transmembrane tran
161	33	54.1	1260	2	A86323	protein F14D16.3 [	234	32	52.5	527	2	D87318	conserved hypotet
162	33	54.1	1399	2	A99720	RhaC core protein	235	32	52.5	532	2	T52442	hypothetical prote
163	33	54.1	1401	2	T02255	probable ubiquitou	236	32	52.5	538	2	I68093	PRR2 delta - human
164	33	54.1	1558	2	T29233	hypothetical prote	237	32	52.5	606	2	T13152	WDRI protein - hum
165	33	54.1	2133	2	T30637	hypothetical prote	238	32	52.5	607	1	ABXL72	74K albumin precur
166	33	54.1	2514	1	MNWV82	nonstructural poly	239	32	52.5	615	1	ABCHS	serum albumin prec
167	33	54.1	3144	2	A46068	Huntington diseas	240	32	52.5	620	2	T30765	hypothetical prote
168	33	54.1	8243	2	T31307	type I fatty acid	241	32	52.5	624	1	JC4510	pullulanase (EC 3.
169	32.5	53.3	443	2	C83421	hypothetical prote	242	32	52.5	630	2	D77992	hypothetical prote
170	32	52.5	133	2	E30587	outer membrane pro	243	32	52.5	631	2	S72270	alpha-amylase (EC
171	32	52.5	133	2	F30587	outer membrane pro	244	32	52.5	632	2	A95923	hypothetical prote
172	32	52.5	154	2	H83160	hypothetical prote	245	32	52.5	678	2	A71287	probable cytoplasm
173	32	52.5	155	2	C60333	outer membrane pro	246	32	52.5	758	2	I64084	hemoglobin recepto
174	32	52.5	167	2	C64456	hypothetical prote	247	32	52.5	770	2	A49283	outer layer protei
175	32	52.5	168	1	S19361	hypothetical prote	248	32	52.5	773	2	T01539	hypothetical prote





395	31	50.8	686	2	A59348	formate dehydrogen	468	30	49.2	170	2	T50215	probable peptide m
396	31	50.8	687	2	A41905	ferric vibriobacti	469	30	49.2	183	2	A81821	DNA-3-methyladenin
397	31	50.8	692	2	T47493	hypothetical prote	470	30	49.2	183	2	C81057	DNA-3-methyladenin
398	31	50.8	716	2	E89998	conserved hypothet	471	30	49.2	184	2	S34250	27.2K structural p
399	31	50.8	725	1	E64211	virulence-associat	472	30	49.2	187	2	S62511	probable peptide m
400	31	50.8	769	1	IJHJLM	leukocyte adhesio	473	30	49.2	204	2	T19249	hypothetical prote
401	31	50.8	769	1	JC1121	leukocyte adhesio	474	30	49.2	206	2	JL0059	H-2 class I histoc
402	31	50.8	770	2	S04847	leukocyte adhesio	475	30	49.2	206	2	JL0058	H-2 class I histoc
403	31	50.8	771	2	A45839	leukocyte adhesio	476	30	49.2	211	2	E71334	hypothetical prote
404	31	50.8	776	2	T29064	hyaluronate lyase	477	30	49.2	216	2	A61259	glycoprotein S - p
405	31	50.8	785	2	T09491	hemagglutinin, pha	478	30	49.2	218	2	T49885	peptide methionine
406	31	50.8	797	2	AH1302	primosomal replica	479	30	49.2	219	2	T17539	glycerophosphoryl
407	31	50.8	797	2	AH1674	primosomal replica	480	30	49.2	221	2	A57296	ribosomal protein
408	31	50.8	807	2	D86350	DNA topoisomerase	481	30	49.2	238	2	B64404	hypothetical prote
409	31	50.8	813	2	A13479	hypothetical prote	482	30	49.2	249	2	T19088	hypothetical prote
410	31	50.8	862	2	F75116	endoglucanase I (E	483	30	49.2	249	2	T31837	hypothetical prote
411	31	50.8	879	2	A47704	valyl-tRNA synthet	484	30	49.2	253	2	A36564	hemoglobin linker
412	31	50.8	882	2	AB1631	hypothetical prote	485	30	49.2	261	2	A86763	cobryic acid synth
413	31	50.8	899	2	I38153	gene ret11 protein	486	30	49.2	261	2	T07630	expansin 1 - tomat
414	31	50.8	937	2	A45082	neurotrophic recep	487	30	49.2	299	2	T26789	hypothetical prote
415	31	50.8	940	2	S19702	fibronectin-bindin	488	30	49.2	299	2	S54087	probable membrane
416	31	50.8	961	2	G90053	hypothetical prote	489	30	49.2	307	2	B72031	lipic acid synth
417	31	50.8	987	2	A64474	hypothetical prote	490	30	49.2	307	2	F86594	lipocate synthetase
418	31	50.8	1038	2	H90053	hypothetical prote	491	30	49.2	309	2	A38395	mast cell carboxyp
419	31	50.8	1050	2	S45636	natriuretic-peptid	492	30	49.2	314	2	A00220	flagellar protein
420	31	50.8	1051	2	T18351	lmp1 protein - Myc	493	30	49.2	314	2	AF3193	MHC nonclassical c
421	31	50.8	1097	2	T40678	hypothetical prote	494	30	49.2	319	2	B35090	hypothetical prote
422	31	50.8	1114	1	S05582	protein-tyrosine k	495	30	49.2	319	2	D64303	MHC class I histoc
423	31	50.8	1120	2	H71664	transcription-repa	496	30	49.2	326	2	A32273	MHC class I histoc
424	31	50.8	1127	2	T25804	hypothetical prote	497	30	49.2	328	2	I54414	MHC H-2K transplan
425	31	50.8	1300	2	S73679	probable lipoprote	498	30	49.2	328	2	S51458	hypothetical prote
426	31	50.8	1301	2	S18118	alpha-amylase - Al	499	30	49.2	330	2	AD2258	hypothetical prote
427	31	50.8	1302	1	JC6009	surface-located me	500	30	49.2	330	2	AC2240	hypothetical prote
428	31	50.8	1347	2	T02214	ubiquitous TPR mot	501	30	49.2	332	2	G96840	hypothetical prote
429	31	50.8	1353	2	T26301	hypothetical prote	502	30	49.2	347	2	T32768	hypothetical prote
430	31	50.8	1365	2	T30822	lmp1 protein - Myc	503	30	49.2	347	2	E84898	hypothetical prote
431	31	50.8	1475	2	T29809	hypothetical prote	504	30	49.2	348	2	A38984	C4-dicarboxylate t
432	31	50.8	1513	1	RNR222	DNA-directed RNA p	505	30	49.2	351	2	C91182	probable fibrial
433	31	50.8	1638	2	T25352	hypothetical prote	506	30	49.2	351	2	T44428	probable gonococca
434	31	50.8	1722	2	E89753	protein p11C7.4 li	507	30	49.2	351	2	G86028	probable fibrial
435	31	50.8	1827	1	A23945	sucrose alpha-gluc	508	30	49.2	351	2	A23351	serendipity (ary)
436	31	50.8	1854	2	S36859	c1pA protein - Clo	509	30	49.2	353	2	S25536	MHC class I histoc
437	31	50.8	2215	2	T16871	hypothetical prote	510	30	49.2	356	2	A21198	H-2 class I histoc
438	31	50.8	2481	2	A43908	fibronectin - Afri	511	30	49.2	356	2	T18590	hypothetical prote
439	31	50.8	2581	2	AP2545	hypothetical prote	512	30	49.2	358	2	S75659	gene At103 protein
440	31	50.8	4544	1	S02352	alpha-2-macroglobu	513	30	49.2	360	2	I54554	MHC class I histoc
441	31	50.8	4545	1	S25111	alpha-2-macroglobu	514	30	49.2	362	1	HLMSDB	MHC class I histoc
442	31	50.8	4976	2	T14165	peptide synthetase	515	30	49.2	362	2	C60854	MHC class I histoc
443	30.5	50.0	454	2	S75741	hypothetical prote	516	30	49.2	362	2	G08854	MHC class I histoc
444	30.5	50.0	523	2	T36968	probable phytoene	517	30	49.2	364	1	B60854	protein-tyrosine-p
445	30.5	50.0	2183	1	GNBY84	genome polypotein	518	30	49.2	364	1	S31304	MHC class I histoc
446	30	49.2	53	2	S50105	hypothetical 5.8K	519	30	49.2	365	2	I57814	MHC class I histoc
447	30	49.2	55	2	C82814	hypothetical prote	520	30	49.2	367	2	S31651	MHC class I histoc
448	30	49.2	91	2	I54404	T cell antigen rec	521	30	49.2	368	1	HLMSKD	MHC class I Rtl.E
449	30	49.2	98	2	S03497	T-cell receptor be	522	30	49.2	368	2	I49712	MHC class I histoc
450	30	49.2	113	2	I38312	T-cell receptor be	523	30	49.2	368	2	I49713	MHC class I histoc
451	30	49.2	116	2	T24843	hypothetical prote	524	30	49.2	368	2	I68705	protein kinase (EC
452	30	49.2	116	2	T16332	hypothetical prote	525	30	49.2	369	1	HLMSKB	protein kinase (EC
453	30	49.2	116	2	H90966	hypothetical prote	526	30	49.2	369	1	HLMSKB	protein kinase (EC
454	30	49.2	119	2	H90513	hypothetical prote	527	30	49.2	370	2	I54531	protein kinase (EC
455	30	49.2	129	1	QCQVW4	AC2 protein - abut	528	30	49.2	373	2	I69008	MHC class I Rtl.E
456	30	49.2	129	2	JG0019	flagellar basal-bo	529	30	49.2	377	2	A45851	MHC class I histoc
457	30	49.2	134	2	A30563	T-cell receptor be	530	30	49.2	379	1	OKHUR1	protein kinase (EC
458	30	49.2	137	2	T15522	hypothetical prote	531	30	49.2	379	1	OKB01R	protein kinase (EC
459	30	49.2	140	2	S48562	hypothetical prote	532	30	49.2	380	1	OKPGLR	protein kinase (EC
460	30	49.2	142	2	S19245	Ig heavy chain pre	533	30	49.2	381	1	A60669	protein kinase (EC
461	30	49.2	142	4	S13768	MHC class I histoc	534	30	49.2	381	1	OKHULR	protein kinase (EC
462	30	49.2	145	2	S65983	yybN protein - Bac	535	30	49.2	381	1	OKMSR1	protein kinase (EC
463	30	49.2	148	2	AC1897	hypothetical prote	536	30	49.2	381	1	OKRT1R	protein kinase (EC
464	30	49.2	161	2	S74336	hypothetical prote	537	30	49.2	385	2	T18180	proline-rich prote
465	30	49.2	161	2	E89812	conserved hypothet	538	30	49.2	392	2	A60777	keratin 2, type I,
466	30	49.2	164	2	T16321	hypothetical prote	539	30	49.2	395	1	A25887	calsequestrin prec
467	30	49.2	165	2	S22511	hypothetical prote	540	30	49.2	395	1	A46345	gene III protein -

541	30	49.2	406	1	A31050	calsequestrin prec	614	30	49.2	640	1	ALASGR	glucan 1,4-alpha-g
542	30	49.2	407	2	S50870	cell division cont	615	30	49.2	640	2	A29166	glucan 1,4-alpha-g
543	30	49.2	408	2	T24114	hypothetical prote	616	30	49.2	640	2	A64065	hypothetical prote
544	30	49.2	409	1	JQ1396	calsequestrin prec	617	30	49.2	647	2	A49218	hemagglutinin homo
545	30	49.2	410	1	A28071	calsequestrin prec	618	30	49.2	648	2	T28325	ORF MSV164 probabl
546	30	49.2	414	2	T50010	hypothetical prote	619	30	49.2	649	2	S48752	major surface prot
547	30	49.2	420	1	S22418	calsequestrin prec	620	30	49.2	655	1	A46688	hepatocyte growth
548	30	49.2	421	2	S67952	probable membrane	621	30	49.2	667	2	AD0295	probable formyl tr
549	30	49.2	423	2	S38953	carboxypeptidase D	622	30	49.2	677	2	G84514	hypothetical prote
550	30	49.2	429	2	E75107	thiamin biosynthes	623	30	49.2	685	2	S64158	hypothetical prote
551	30	49.2	431	2	F71600	variant-specific s	624	30	49.2	696	2	T46111	probable transposa
552	30	49.2	438	2	A64428	thiamin biosynthes	625	30	49.2	698	1	TVFVLV	protein-tyrosine k
553	30	49.2	441	2	T11715	probable acetylorn	626	30	49.2	704	2	JC7881	ceramidase (EC 3.5
554	30	49.2	442	2	T42091	acetylornithine tr	627	30	49.2	726	1	S73915	virulence-associat
555	30	49.2	448	2	T45710	H-protein promoter	628	30	49.2	742	1	S58691	kinesin-related pr
556	30	49.2	448	2	C86447	F5D14.3 protein -	629	30	49.2	808	2	A82877	hypothetical prote
557	30	49.2	451	1	JC6180	stearoyl-CoA 9-des	630	30	49.2	826	2	B36203	iron-responsive el
558	30	49.2	452	2	S10840	gene ND4L intron p	631	30	49.2	861	2	T14945	primase - human he
559	30	49.2	457	2	S63608	glutamate dehydrog	632	30	49.2	875	2	T33434	hypothetical prote
560	30	49.2	463	2	G84901	hypothetical prote	633	30	49.2	877	2	H64708	iron-regulated out
561	30	49.2	464	2	AG2879	two component sens	634	30	49.2	879	2	E71811	probable iron-regu
562	30	49.2	464	2	A97656	probable two-compo	635	30	49.2	880	1	SYSVS	valine-tRNA ligase
563	30	49.2	468	1	P2ML51	L2 protein - human	636	30	49.2	902	2	T47966	hypothetical prote
564	30	49.2	470	2	H69296	acetyl-CoA decarbo	637	30	49.2	903	2	B84548	hypothetical prote
565	30	49.2	480	2	AB0148	catalase (EC 1.11.	638	30	49.2	907	2	T27317	hypothetical prote
566	30	49.2	480	2	S72668	oviduct-specific p	639	30	49.2	916	2	F71962	hypothetical prote
567	30	49.2	480	2	JC7866	endo-1,6-beta-D-gl	640	30	49.2	925	2	T00781	hypothetical prote
568	30	49.2	484	2	E64432	spore coat polysac	641	30	49.2	967	2	G86229	beta galactosidase
569	30	49.2	494	2	S57539	probable membrane	642	30	49.2	938	2	T42241	hypothetical prote
570	30	49.2	506	2	AE2475	polyketide synthas	643	30	49.2	941	1	TVMVMD	protein-tyrosine k
571	30	49.2	506	2	B52210	hetM protein - Ana	644	30	49.2	952	2	B57238	iron-responsive el
572	30	49.2	507	2	T38653	trp-asp repeat pro	645	30	49.2	963	2	A57238	iron-responsive el
573	30	49.2	510	2	E37307	probable cardiolip	646	30	49.2	967	2	G86229	hypothetical prote
574	30	49.2	513	2	T38770	alpha-amylase a pr	647	30	49.2	982	2	A97210	beta galactosidase
575	30	49.2	521	2	T34583	probable periplasm	648	30	49.2	985	2	E69850	formate dehydrogen
576	30	49.2	521	2	E82377	ABC transporter, p	649	30	49.2	1001	2	S66704	hypothetical prote
577	30	49.2	522	2	B82024	peptide methionine	650	30	49.2	1004	2	T71490	probable exodeoxyr
578	30	49.2	522	2	G31243	peptide methionine	651	30	49.2	1011	2	T17430	col protein - Neur
579	30	49.2	530	2	JN0597	calnexin-like prot	652	30	49.2	1018	2	A32192	fibronectin-bindin
580	30	49.2	531	2	B88109	protein T24E12.5 [	653	30	49.2	1025	2	G81751	exodeoxyribonuclea
581	30	49.2	532	2	T49873	calnexin homolog -	654	30	49.2	1042	2	G64514	type I restriction
582	30	49.2	534	2	A29776	glucan 1,4-alpha-g	655	30	49.2	1077	2	T41146	probable cyteine-
583	30	49.2	537	2	T48599	hypothetical prote	656	30	49.2	1099	2	I46497	bumetanide-sensiti
584	30	49.2	540	2	T10892	probable calnexin	657	30	49.2	1099	2	I46498	bumetanide-sensiti
585	30	49.2	541	2	T46423	hypothetical prote	658	30	49.2	1099	2	I46496	bumetanide-sensiti
586	30	49.2	546	2	T06415	calnexin - soybean	659	30	49.2	1155	2	S64365	GDP/GTP exchange p
587	30	49.2	547	2	S54513	probable membrane	660	30	49.2	1174	1	HJBYDH	helicase (EC 3.6.1
588	30	49.2	548	2	A70980	probable accB5 pro	661	30	49.2	1180	1	NCECX5	exodeoxyribonuclea
589	30	49.2	555	1	B69862	conserved hypothet	662	30	49.2	1180	2	G85933	DNA helicase RecB
590	30	49.2	558	2	S68447	origin recognition	663	30	49.2	1180	2	E91088	DNA helicase RecB
591	30	49.2	561	2	E30891	hypothetical prote	664	30	49.2	1181	2	AB0865	exonuclease V chai
592	30	49.2	561	2	D85726	hypothetical prote	665	30	49.2	1205	2	T18517	procollagen N-endo
593	30	49.2	561	2	C54903	hypothetical prote	666	30	49.2	1223	1	TVCHLV	epidermal growth f
594	30	49.2	566	2	T33042	probable ABC-type	667	30	49.2	1224	2	T00059	hypothetical prote
595	30	49.2	567	2	G90537	hypothetical prote	668	30	49.2	1254	1	JQ1978	structural polypro
596	30	49.2	571	2	G83226	lipoprotein (limpor	669	30	49.2	1254	1	VHVVE	structural polypro
597	30	49.2	572	2	S77249	hypothetical prote	670	30	49.2	1254	1	VHVVT	structural polypro
598	30	49.2	586	2	T19075	hypothetical prote	671	30	49.2	1255	1	B44213	structural polypro
599	30	49.2	587	2	A60367	transforming prote	672	30	49.2	1255	1	D44213	structural polypro
600	30	49.2	595	2	T06412	probable chaperoni	673	30	49.2	1289	2	T31344	structural polypro
601	30	49.2	602	1	JQ0806	hydrogenase (EC 1.	674	30	49.2	1328	2	T43060	agrin - electric r
602	30	49.2	604	1	TVYUH	protein-tyrosine k	675	30	49.2	1334	2	AB1775	hypothetical prote
603	30	49.2	608	2	S65298	dicarboxylic amino	676	30	49.2	1348	2	S27812	probable epidermal
604	30	49.2	617	2	S08317	hydrogenase (EC 1.	677	30	49.2	1348	2	A43917	probable epidermal
605	30	49.2	619	2	E83635	hypothetical prote	678	30	49.2	1356	2	F84486	probable retrolem
606	30	49.2	620	2	T39074	yeast chs5 homolog	679	30	49.2	1363	2	T47492	copla-like polypro
607	30	49.2	636	2	E30943	probable enzyme li	680	30	49.2	1403	1	A47328	natural killer cel
608	30	49.2	636	2	AB5792	probable enzyme Z2	681	30	49.2	1447	1	VGHEE3	E2 glycoprotein pr
609	30	49.2	636	2	H64941	Probable ATP-depen	682	30	49.2	1447	1	VGHEE2	E2 glycoprotein pr
610	30	49.2	637	2	F96900	DNA gyrase (topois	683	30	49.2	1449	1	VGHFS	E2 glycoprotein pr
611	30	49.2	637	2	T46555	DNA topoisomerase	684	30	49.2	1449	1	VGHFS	E2 glycoprotein pr
612	30	49.2	639	2	JQ0607	glucan 1,4-alpha-g	685	30	49.2	1449	2	S47423	E2 glycoprotein pr
613	30	49.2	639	2	JT0479	glucan 1,4-alpha-g	686	30	49.2	1451	1	JQ1719	E2 glycoprotein pr

687	30	49.2	1452	1	VGIH79	E2 glycoprotein pr	760	29	47.5	268	2	D86270	F21P23.14 protein
688	30	49.2	1453	2	S41453	spike protein - ca	761	29	47.5	268	2	T00827	hypothetical prote
689	30	49.2	1492	2	T18560	DNA-directed DNA p	762	29	47.5	269	2	AE3455	arginine-binding p
690	30	49.2	1538	2	T29095	cardiac muscle fac	763	29	47.5	271	2	F64349	hypothetical prote
691	30	49.2	1680	2	T41628	probable transcrip	764	29	47.5	271	2	S12560	transcription fact
692	30	49.2	1785	2	A45548	major merozoite su	765	29	47.5	272	2	G71618	merozoite surface
693	30	49.2	1856	2	C95008	immunoglobulin A1	766	29	47.5	274	2	B89007	protein F59B1.5 [l
694	30	49.2	2195	2	T34264	hypothetical prote	767	29	47.5	276	2	I40453	licheninase (EC 3.
695	30	49.2	2287	2	T21312	hypothetical prote	768	29	47.5	285	2	C97279	thioredoxin reduct
696	30	49.2	2496	2	A71616	secreted protein p	769	29	47.5	285	2	B82842	spermidine syntha
697	30	49.2	5825	2	T12117	polyprotein - fava	770	29	47.5	287	2	T40138	probable ribosom
698	30	49.2	6359	2	T13679	bacitracin synthe	771	29	47.5	289	2	C69801	L-alanoyl-D-glutam
699	29.5	48.4	296	2	T22710	hypothetical prote	772	29	47.5	289	2	A11448	L-alanoyl-D-glutam
700	29.5	48.4	435	2	C86340	protein F2D10.28 [	773	29	47.5	291	2	H70709	thioredoxin reduct
701	29.5	48.4	648	2	AF0609	conserved hypotet	774	29	47.5	293	1	DDBP32	helix-destabilizin
702	29.5	48.4	800	1	S31575	interleukin-4 rece	775	29	47.5	293	1	DDBP36	helix-destabilizin
703	29.5	48.4	899	2	H87513	hypothetical prote	776	29	47.5	294	2	G83962	hypothetical prote
704	29	47.5	29	2	B54197	70K thyroid autoan	777	29	47.5	295	2	F95997	probable transcrip
705	29	47.5	62	2	C86560	conserved hypotet	778	29	47.5	295	4	S36174	RNA binding protei
706	29	47.5	62	2	B72064	conserved hypotet	779	29	47.5	296	2	B82866	conjugal transfer
707	29	47.5	71	2	T00008	copy number contro	780	29	47.5	296	2	C82605	conjugal transfer
708	29	47.5	71	2	T07454	hypothetical prote	781	29	47.5	299	2	B90111	DNA-directed RNA p
709	29	47.5	84	2	JN0469	85K WRK-20 recogni	782	29	47.5	301	1	DDBP34	helix-destabilizin
710	29	47.5	96	2	T13118	secondary immunity	783	29	47.5	305	1	PRYNB	Na+/K+-exchangin
711	29	47.5	96	2	AG0229	conserved hypotet	784	29	47.5	307	2	E89075	protein K04A8.2 [l
712	29	47.5	100	2	A46503	beta-integrin - ra	785	29	47.5	310	2	S36581	E2 protein - human
713	29	47.5	107	2	S08521	hypothetical 11.7K	786	29	47.5	310	2	AB2235	hypothetical prote
714	29	47.5	115	2	B84969	flagellar basal-bo	787	29	47.5	311	2	T49912	hypothetical prote
715	29	47.5	119	2	AF2509	hypothetical prote	788	29	47.5	312	2	S04281	psbA intron 2 prot
716	29	47.5	129	2	H70727	hypothetical prote	789	29	47.5	314	2	B90334	hypothetical prote
717	29	47.5	140	2	AC0512	probable secreted	790	29	47.5	323	2	S01895	T-cell receptor ga
718	29	47.5	141	2	H85060	hypothetical prote	791	29	47.5	323	2	T46285	hypothetical prote
719	29	47.5	143	2	AH3060	hypothetical prote	792	29	47.5	324	2	JC2395	Fas antigen precu
720	29	47.5	158	2	T49567	related to attachm	793	29	47.5	332	2	T19150	hypothetical prote
721	29	47.5	163	2	T71635	hypothetical prote	794	29	47.5	334	2	H95307	arginine deiminase
722	29	47.5	167	2	H98225	hypothetical prote	795	29	47.5	335	2	A85913	probable enzyme X3
723	29	47.5	168	2	S71562	drought-induced pr	796	29	47.5	335	2	G91068	probable enzyme EC
724	29	47.5	176	2	G81269	probable acetyltra	797	29	47.5	336	2	T10733	cinnamoyl-CoA redu
725	29	47.5	177	2	S47348	histone H1.0 - rat	798	29	47.5	336	2	A89772	hypothetical prote
726	29	47.5	180	2	A56113	surface antigen FU	799	29	47.5	337	2	A83403	ribose operon repr
727	29	47.5	184	2	G88449	protein F54D8.5 [l	800	29	47.5	339	2	H97107	uncharacterized pr
728	29	47.5	186	2	G64397	hypothetical prote	801	29	47.5	343	2	G71979	probable type II D
729	29	47.5	188	2	A29867	hypothetical 20K p	802	29	47.5	351	2	T50105	hypothetical prote
730	29	47.5	194	2	HSU010	histone H1.0 - hum	803	29	47.5	351	2	G84128	hypothetical prote
731	29	47.5	194	2	I49150	histone H1.0 - mou	804	29	47.5	352	2	AC3306	periplasmic dipept
732	29	47.5	196	1	HSX05A	histone H5A - Afri	805	29	47.5	354	2	S27014	GTP-binding regula
733	29	47.5	196	2	A30484	histone H5B - Afri	806	29	47.5	355	2	E81415	DNA-directed DNA p
734	29	47.5	210	2	B65045	hypothetical prote	807	29	47.5	356	2	S61061	hypothetical prote
735	29	47.5	211	2	T24760	hypothetical prote	808	29	47.5	361	2	I48160	MHC class I protei
736	29	47.5	214	2	B46244	insulin-like growt	809	29	47.5	363	2	S61970	hypothetical prote
737	29	47.5	217	2	T19005	hypothetical prote	810	29	47.5	363	2	S49539	Cood protein precu
738	29	47.5	220	2	AE3221	conserved hypotet	811	29	47.5	365	2	C97735	hypothetical prote
739	29	47.5	224	2	E96982	ortholog yrbG, yet	812	29	47.5	369	2	S63464	hypothetical prote
740	29	47.5	224	2	G72746	hypothetical prote	813	29	47.5	371	2	T03025	mitosis-specific c
741	29	47.5	224	2	B72710	hypothetical prote	814	29	47.5	371	2	D97042	hypothetical prote
742	29	47.5	228	2	I58170	LEKK-7 precursor -	815	29	47.5	373	2	S48451	probable membrane
743	29	47.5	234	2	T26363	hypothetical prote	816	29	47.5	373	2	AH0855	lipoprotein NlpD p
744	29	47.5	240	2	T38050	hypothetical prote	817	29	47.5	376	2	G71925	cag island protein
745	29	47.5	241	2	T23428	hypothetical prote	818	29	47.5	377	2	T40024	probable cytochrom
746	29	47.5	246	2	T30490	hypothetical prote	819	29	47.5	379	2	D91078	probable lipoprote
747	29	47.5	246	2	T37473	transcription regu	820	29	47.5	379	2	B85923	lipoprotein (impor
748	29	47.5	247	2	T21406	hypothetical prote	821	29	47.5	379	2	B55522	lipoprotein D prec
749	29	47.5	250	2	S61556	Rieske iron-sulfur	822	29	47.5	381	2	S37170	repB protein - Lac
750	29	47.5	251	2	T25121	hypothetical prote	823	29	47.5	383	2	B86272	protein F16A14.12
751	29	47.5	255	1	E71090	probable lactam ut	824	29	47.5	384	2	A29130	calreticulin (clon
752	29	47.5	255	2	H75074	lactam utilization	825	29	47.5	384	2	AB2024	sulfolipid biosynt
753	29	47.5	257	2	H84157	essential for sigm	826	29	47.5	388	2	S40057	repA protein - Lac
754	29	47.5	257	2	T21029	hypothetical prote	827	29	47.5	388	2	T22795	hypothetical prote
755	29	47.5	260	2	H97357	stage 0 sporulatio	828	29	47.5	392	2	S37902	hypothetical prote
756	29	47.5	260	2	S47724	hypothetical 29.5K	829	29	47.5	401	2	T17515	hypothetical prote
757	29	47.5	261	2	S20610	calpastatin - mous	830	29	47.5	402	2	E64440	hypothetical prote
758	29	47.5	262	2	T43753	probable COI intro	831	29	47.5	409	2	D83326	probable acyl-CoA
759	29	47.5	264	2	AE2274	hypothetical prote	832	29	47.5	411	2	S29129	calreticulin precu

833	29	47.5	414	2	D96838	unknown protein T2	906	29	47.5	575	2	S75395	probable glutamine
834	29	47.5	416	1	S06763	calreticulin precu	907	29	47.5	579	2	B86158	F22p16.20 protein
835	29	47.5	416	2	JH0819	calreticulin precu	908	29	47.5	582	2	H37017	ATP-dependent Zn p
836	29	47.5	417	1	A37047	calreticulin precu	909	29	47.5	582	2	A70755	hypothetical prote
837	29	47.5	418	1	A34154	calreticulin precu	910	29	47.5	585	2	C82157	hypothetical prote
838	29	47.5	419	2	S71343	calreticulin precu	911	29	47.5	586	2	S63386	HOL1 protein - yea
839	29	47.5	419	2	F72866	alkaline exonuclea	912	29	47.5	589	2	JG7520	endo-1,6-alpha-D-m
840	29	47.5	419	2	E86292	FYH2.15 protein -	913	29	47.5	598	2	T28238	ORF MSV077 hypothe
841	29	47.5	420	2	T41870	ALK-EXO ofr133 - B	914	29	47.5	599	2	A11025	hypothetical prote
842	29	47.5	421	2	T25383	hypothetical prote	915	29	47.5	602	2	T13219	major capsid prote
843	29	47.5	423	2	T50923	acetylornithine tr	916	29	47.5	603	2	S15074	calpastatin - rat
844	29	47.5	423	2	S57103	hypothetical prote	917	29	47.5	607	2	A47757	retrovirus-related
845	29	47.5	425	2	E64657	probable glucose-6	918	29	47.5	609	2	A30894	70K thyroid autoan
846	29	47.5	425	2	T34872	hypothetical prote	919	29	47.5	610	2	A84417	hypothetical prote
847	29	47.5	427	2	T19004	hypothetical prote	920	29	47.5	614	2	T19963	hypothetical prote
848	29	47.5	429	2	S09141	ND4L intron 2 prot	921	29	47.5	619	2	A60646	transforming prote
849	29	47.5	429	2	S45459	TM34 protein - ye	922	29	47.5	624	2	PC6003	surface membrane p
850	29	47.5	432	1	S18932	u-plasminogen acti	923	29	47.5	624	2	B83386	hypothetical prote
851	29	47.5	433	1	JN0560	u-plasminogen acti	924	29	47.5	634	2	AG0252	probable DEAD box
852	29	47.5	433	2	A81793	deoxyribodipyrimid	925	29	47.5	636	2	AF0725	probable ATP-depen
853	29	47.5	436	1	S49458	diphosphate-fructo	926	29	47.5	644	2	S50429	dnak-type molecula
854	29	47.5	437	2	F59099	hypothetical prote	927	29	47.5	654	2	T10772	calpastatin - rat
855	29	47.5	442	1	UKPG	u-plasminogen acti	928	29	47.5	656	1	A34890	histidine decarbox
856	29	47.5	449	1	KIECD3	aspartate kinase (	929	29	47.5	658	2	T03416	traG protein - Agr
857	29	47.5	449	2	G91254	lysine sensitive a	930	29	47.5	658	2	AB3243	conjugal transfer
858	29	47.5	449	2	C86095	aspartokinase III,	931	29	47.5	662	1	S12989	histidine decarbox
859	29	47.5	449	2	T08309	conserved hypothet	932	29	47.5	664	2	A97222	membrane associate
860	29	47.5	455	2	T38275	hypothetical prote	933	29	47.5	668	2	I39902	penicillin-binding
861	29	47.5	456	2	T45610	proanthranilate N-	934	29	47.5	682	1	RNRZC1	DNA-directed RNA p
862	29	47.5	457	2	C86669	amino acid permeas	935	29	47.5	684	2	T33785	hypothetical prote
863	29	47.5	458	2	A84487	probable replicati	936	29	47.5	688	2	S85241	hypothetical prote
864	29	47.5	463	2	T16503	hypothetical prote	937	29	47.5	705	2	C75118	dipeptide abc tran
865	29	47.5	464	2	B57720	kinate receptor b	938	29	47.5	719	2	AB1358	hypothetical prote
866	29	47.5	470	1	NMIVW8	exo-alpha-sialidas	939	29	47.5	732	2	A43315	BTS domain protein
867	29	47.5	470	1	NMIV9	exo-alpha-sialidas	940	29	47.5	736	2	A99279	hypothetical prote
868	29	47.5	473	2	C84312	glycine dehydrogen	941	29	47.5	736	2	E71414	hypothetical prote
869	29	47.5	474	2	S30168	mercury(II) reduct	942	29	47.5	750	1	COZPME	mei2 protein - fis
870	29	47.5	474	2	A35732	protective protein	943	29	47.5	751	2	I48748	semaphorin E - mou
871	29	47.5	475	2	B75024	glutamate synthase	944	29	47.5	754	2	T16182	hypothetical prote
872	29	47.5	475	2	F70144	carboxyl-terminal	945	29	47.5	754	2	AH3004	vgrG protein limpo
873	29	47.5	476	2	D11200	probable glutamate	946	29	47.5	759	2	S62067	TV1 enhancer activ
874	29	47.5	481	2	AB1527	hypothetical prote	947	29	47.5	767	2	S41479	DNA-binding protei
875	29	47.5	489	2	S67802	hypothetical prote	948	29	47.5	775	2	T45238	probable transfera
876	29	47.5	489	2	S76768	hypothetical prote	949	29	47.5	797	2	D71621	hypothetical prote
877	29	47.5	492	2	T26502	hypothetical prote	950	29	47.5	798	2	A40526	integrin beta-7 ch
878	29	47.5	492	2	S75049	hypothetical prote	951	29	47.5	806	2	A46271	integrin beta-7 ch
879	29	47.5	502	2	AE3569	fructuronate reduc	952	29	47.5	811	2	S77577	endo-alpha-sialida
880	29	47.5	502	2	T17416	probable alkylali	953	29	47.5	816	2	G84952	aspartate kinase (
881	29	47.5	505	2	C82216	probable fumarate	954	29	47.5	816	2	B98196	hypothetical prote
882	29	47.5	510	2	H84887	probable pectinest	955	29	47.5	816	2	AH3090	VgrG protein limpo
883	29	47.5	514	2	T26501	hypothetical prote	956	29	47.5	828	2	AH2443	hypothetical prote
884	29	47.5	517	2	A49776	xylan 1,4-beta-xy	957	29	47.5	836	2	AF3233	conserved hypothet
885	29	47.5	521	2	H82918	formate-tetrahydro	958	29	47.5	845	2	T52518	related to cytosin
886	29	47.5	526	2	A86274	FYA19.15 protein -	959	29	47.5	850	2	S45553	SIN3 protein-bindi
887	29	47.5	527	2	F83319	probable thiosulfa	960	29	47.5	851	2	AD1427	internain, probab
888	29	47.5	529	2	C82314	methyl-accepting c	961	29	47.5	853	2	H70470	hypothetical prote
889	29	47.5	529	2	S35306	phytoene dehydroge	962	29	47.5	855	2	T38754	beta-transducin -
890	29	47.5	531	2	T11596	hypothetical prote	963	29	47.5	857	2	T05172	hypothetical prote
891	29	47.5	540	1	TVFVEB	protein-tyrosine k	964	29	47.5	859	1	VCLJCT	env polyprotein pr
892	29	47.5	540	2	D86737	malolactac enzyme	965	29	47.5	863	2	T38016	importin beta-1 su
893	29	47.5	540	2	S38728	malolactac enzyme	966	29	47.5	863	2	AH2296	hypothetical prote
894	29	47.5	540	2	B44776	protein-tyrosine k	967	29	47.5	879	2	S49910	chloroplast outer
895	29	47.5	544	2	S35745	protein-tyrosine k	968	29	47.5	882	2	A39030	androgen-binding p
896	29	47.5	545	2	S00727	kinase-related tra	969	29	47.5	896	2	S26984	probable DNA-direc
897	29	47.5	546	1	VGN2RK	cell fusion glycop	970	29	47.5	903	2	T09143	alpha-glucosidase
898	29	47.5	546	2	S47305	gene F protein - r	971	29	47.5	903	2	A10015	maltoose regulon po
899	29	47.5	547	2	T25478	hypothetical prote	972	29	47.5	913	2	JC5463	alpha-glucosidase
900	29	47.5	556	2	AD11394	arginyl tRNA synth	973	29	47.5	914	2	S46593	finger protein AZP
901	29	47.5	556	2	AD1770	arginyl tRNA synth	974	29	47.5	931	2	D86222	protein F7G19.9 li
902	29	47.5	556	2	S22634	sphingomyelin phos	975	29	47.5	938	2	A39160	transcription acti
903	29	47.5	558	2	C71609	hypothetical prote	976	29	47.5	940	2	T41992	hypothetical prote
904	29	47.5	570	2	AD2292	hypothetical prote	977	29	47.5	941	2	I40772	hypothetical prote
905	29	47.5	574	2	T33794	hypothetical prote	978	29	47.5	946	2	F81361	probable cell divi

979 29 47.5 955 2 T39765 probable nuclear m  
980 29 47.5 966 2 T30017 hypothetical prote  
981 29 47.5 969 2 T23256 hypothetical prote  
982 29 47.5 978 1 RGBYI3 regulatory protein  
983 29 47.5 984 2 H90029 hypothetical prote  
984 29 47.5 989 2 C83035 hypothetical prote  
985 29 47.5 996 2 JE0237 apolipoprotein E r  
986 29 47.5 1004 1 S55353 probable copper-tr  
987 29 47.5 1025 1 JC1266 beta-galactosidase  
988 29 47.5 1047 2 T25782 hypothetical prote  
989 29 47.5 1073 2 T01955 hypothetical prote  
990 29 47.5 1082 1 RNEG2B DNA-directed RNA p  
991 29 47.5 1093 2 A31758 phosphorylase kina  
992 29 47.5 1093 2 B40793 phosphorylase kina  
993 29 47.5 1097 2 JQ0301 hypothetical 127K  
994 29 47.5 1141 2 A44093 cGMP-inhibited cAM  
995 29 47.5 1172 2 A42587 thrombospondin 2 p  
996 29 47.5 1178 1 A39804 thrombospondin pre  
997 29 47.5 1208 2 T39068 coiled coil protei  
998 29 47.5 1254 1 JQ1979 structural polypro  
999 29 47.5 1268 2 T18955 hypothetical prote  
1000 29 47.5 1280 2 T34357 hypothetical prote

## ALIGNMENTS

RESULT 1  
A83547  
probable aconitate hydratase PA0794 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2004  
C:Accession: A83547  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: A83547  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-868 <STO>  
A:Cross-references: UNIPROT:Q915E4; UNIPARC:UPI00000C5157; GB:AE004514; GB:AE004091; NID  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0794  
C:Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 70.5%; Score 43; DB 2; Length 868;  
Best Local Similarity 87.5%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8  
|||||  
Db 411 CTNTSNPR 418  
RESULT 2  
B82213  
aconitase hydratase 1 VC1338 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
C:Species: Vibrio cholerae  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 31-Dec-2004  
C:Accession: B82213  
R:Heidelberg, J.P.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, R  
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833; PMID:10952301  
A:Accession: B82213  
A>Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-868 <HEI>  
A:Cross-references: UNIPROT:Q9KSC0; UNIPARC:UPI00000C2F63; GB:AE004213; GB:AE003852; NID:  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC1338  
A:Map position: 1  
C:Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)  
Query Match 70.5%; Score 43; DB 2; Length 868;  
Best Local Similarity 87.5%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8  
|||||  
Db 415 CTNTSNPR 422  
RESULT 3  
C81200  
aconitase hydratase 1 NMB0433 [imported] - Neisseria meningitidis (strain MC58 serogroup  
C:Species: Neisseria meningitidis  
C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 31-Dec-2004  
C:Accession: C81200  
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Massignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755; PMID:10710307  
A:Accession: C81200  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-868 <TET>  
A:Cross-references: UNIPROT:Q9K0X3; UNIPARC:UPI00000C44A8; GB:AE002399; GB:AE002098; NID:  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB0433  
C:Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 70.5%; Score 43; DB 2; Length 868;  
Best Local Similarity 87.5%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8  
|||||  
Db 412 CTNTSNPR 419  
RESULT 4  
H81775  
aconitase hydratase (EC 4.2.1.3) NMA2052 [imported] - Neisseria meningitidis (strain Z2491)  
C:Species: Neisseria meningitidis  
C>Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 31-Dec-2004  
C:Accession: H81775  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell  
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: A81775; MUID:20222556; PMID:10761919  
A:Accession: H81775  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-868 <PAR>  
A:Cross-references: UNIPROT:Q9J705; UNIPARC:UPI00000332C3; GB:AL162758; GB:AL157959; NID:  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: acnA; NMA2052  
C:Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)  
C:Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 70.5%; Score 43; DB 2; Length 868;  
Best Local Similarity 87.5%; Pred. No. 13;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNR 8

Db 412 CTNTSNPR 419

# RESULT 5

S01982

hypothetical protein 87 - phage T5

C:Species: phage T5

C>Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 09-Jul-2004

C:Accession: S01982

R;Kalinan, A.V.; Kryukov, V.M.; Bayev, A.A.

Nucleic Acids Res. 16, 6230, 1998

A>Title: The nucleotide sequence of bacteriophage T5 DNA at the region between early and late genes

A:Reference number: S01982; PMID:88289370; PMID:3267228

A:Accession: S01982

A>Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-87 <KAL>

A:Cross-references: UNIPROT:P13390; UNIPARC:UPI000017A852; EMBL:X07559

C:Genetics:

A:Start codon: GTG

Query Match 67.2%; Score 41; DB 2; Length 87;

Best Local Similarity 70.0%; Pred. No. 3.2;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTNTDNR 10

Db 61 CTNTDNLK 70

# RESULT 6

D86927

dihydrodipicolinate synthase (EC 4.2.1.52) [imported] - Lactococcus lactis subsp. lactis

C:Species: Lactococcus lactis subsp. lactis

C>Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 09-Jul-2004

C:Accession: D86927

R;Bolotin, A.; Winkler, P.; Mager, S.; Jaillon, O.; Malarre, K.; Weissenbach, J.; Ehrlich

Genome Res. 11, 731-753, 2001

A>Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis subsp. lactis

A:Reference number: A86625; PMID:21235186; PMID:11337471

A:Accession: D86827

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-297 <STO>

A:Cross-references: UNIPROT:Q9CF61; UNIPARC:UPI000012854B; GB:AE005176; PID:gi2724628; E

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: dapA

C:Superfamily: dihydrodipicolinate synthase

C:Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 65.6%; Score 40; DB 2; Length 297;

Best Local Similarity 77.8%; Pred. No. 15;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTNTDNR 9

Db 174 CTNTDNLAY 182

# RESULT 7

F97218

glycosyltransferase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004

C:Accession: F97218

R;Nolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

J. Bacteriol. 183, 4823-4838, 2001

A:Reference number: A86625; PMID:21235186; PMID:11337471

A:Accession: D86827

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-297 <STO>

A:Cross-references: UNIPROT:Q9CF61; UNIPARC:UPI000012854B; GB:AE005176; PID:gi2724628; E

A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum

A:Reference number: A96900; PMID:21359325; PMID:21359325

A:Accession: F97218

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-259 <KUR>

A:Cross-references: UNIPROT:Q97FY6; UNIPARC:UPI00000CA596; GB:AE001437; PIDN:AAK80537.1;

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC3588

C:Superfamily: dolichyl-phosphate beta-D-mannosyltransferase

Query Match 63.9%; Score 39; DB 2; Length 259;

Best Local Similarity 66.7%; Pred. No. 20;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTNTDNR 9

Db 103 CDNTDPRY 111

# RESULT 8

T44652

UDP-N-acetylglucosamine-2-epimerase cpso [imported] - Streptococcus agalactiae

C:Species: Streptococcus agalactiae

C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004

C:Accession: T44652

R;Chaffin, D.O.; Yim, H.H.; Beres, S.B.; Sweet, E.S.; Nittayaajarn, A.; Rubens, C.E.

submitted to the EMBL Data Library, June 1999

A:Reference number: Z22821

A:Accession: T44652

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-384 <CHA>

A:Cross-references: UNIPROT:Q9RPB9; UNIPARC:UPI00000B43B7; EMBL:AF163833; PIDN:AAD53075.1

A:Experimental source: strain COH1; serotype III

C:Genetics:

A:Gene: cpso

Query Match 62.3%; Score 38; DB 2; Length 384;

Best Local Similarity 77.8%; Pred. No. 44;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 T44652 10

Db 342 T44652 350

# RESULT 9

T49424

hypothetical protein B17C10.50 [imported] - Neurospora crassa

C:Species: Neurospora crassa

C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 09-Jul-2004

C:Accession: T49424

R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,

submitted to the Protein Sequence Database, May 2000

A:Reference number: Z25022

A:Accession: T49424

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-487 <SCH>

A:Cross-references: UNIPROT:Q9P6E4; UNIPARC:UPI000006AEB4; EMBL:AL355926; GSPDB:GN00116;

A:Experimental source: BAC clone B17C10; strain OR74A

C:Genetics:

A:Gene: NCSP:B17C10.50

A:Map position: 6

A:Introns: 86/1; 150/2

C:Superfamily: Neurospora crassa hypothetical protein B17C10.50

Query Match 62.3%; Score 38; DB 2; Length 487;

Best Local Similarity 85.7%; Pred. No. 56;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;



Qy 1 CTNTDNP 7  
Db 150 CTGTDNP 156

RESULT 10  
T07611  
aconitate hydratase (EC 4.2.1.3) - potato (fragment)  
N;Alternate names: aconitase  
C;Species: Solanum tuberosum (potato)  
C;Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 31-Dec-2004  
C;Accession: T07611  
R;Surpili, M.J.  
submitted to the EMBL Data Library, April 1996  
A;Reference number: Z16049  
A;Accession: T07611  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-616 <SUR>  
A;Cross-references: UNIPROT:O04916; UNIPARC:UPI00001252D0; EMBL:X97012; PIDN:CAA65735.1  
A;Experimental source: cv. Desiree  
C;Function:  
A;Description: reversibly catalyzes the hydration of cis-aconitate to citrate and also to  
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)  
C;Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 62.3%; Score 38; DB 2; Length 616;  
Best Local Similarity 85.7%; Pred. No. 70;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7  
Db 159 CTNTSNP 165

RESULT 11  
G85708  
aconitate hydratase (EC 4.2.1.3) [imported] - Lactococcus lactis subsp. lactis (strain I  
C;Species: Lactococcus lactis subsp. lactis  
C;Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 31-Dec-2004  
C;Accession: G86708  
R;Rolotin, A.; Winkler, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich  
Genome Res. 11, 731-753, 2001  
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s  
A;Reference number: A86625; MUID:21235186; PMID:11337471  
A;Accession: G86708  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-848 <STO>  
A;Cross-references: UNIPROT:Q9CHQ5; UNIPARC:UPI00000C68A0; GB:AE005176; PID:gl2723578; F  
A;Experimental source: strain IL1403  
C;Genetics:  
A;Gene: citB  
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)  
C;Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 62.3%; Score 38; DB 2; Length 848;  
Best Local Similarity 85.7%; Pred. No. 95;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7  
Db 400 CTNTSNP 406

RESULT 12  
C90262  
aconitate hydratase [imported] - Sulfolobus solfataricus  
C;Species: Sulfolobus solfataricus  
C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 31-Dec-2004  
C;Accession: C90262  
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, H

arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A;Description: Sulfolobus solfataricus complete genome.  
A;Reference number: A99139  
A;Accession: C90262  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-855 <KUR>  
A;Cross-references: UNIPROT:Q97242; UNIPARC:UPI0000064353; GB:AE006641; NID:gl3814284; P  
C;Genetics:  
A;Gene: SS01095  
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 855;  
Best Local Similarity 85.7%; Pred. No. 95;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7  
Db 402 CTNTSNP 408

RESULT 13  
E72541  
probable aconitate hydratase APE1618 - Aeropyrum pernix (strain K1)  
C;Species: Aeropyrum pernix  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 31-Dec-2004  
C;Accession: E72541  
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999  
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A;Reference number: A72450; MUID:99310339; PMID:10382966  
A;Accession: E72541  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-870 <KAW>  
A;Cross-references: UNIPROT:Q9VB10; UNIPARC:UPI000005E007; DDBJ:AP0000062; NID:gs105244; I  
A;Experimental source: strain K1  
C;Genetics:  
A;Gene: APE1618  
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 870;  
Best Local Similarity 85.7%; Pred. No. 97;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7  
Db 406 CTNTSNP 412

RESULT 14  
A97854  
aconitate hydratase (EC 4.2.1.3) [imported] - Rickettsia conorii (strain Malish 7)  
C;Species: Rickettsia conorii  
C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 31-Dec-2004  
C;Accession: A97854  
R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro  
Science 293, 2093-2098, 2001  
A;Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A;Reference number: A97700; MUID:21442074; PMID:11557893  
A;Accession: A97854  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-878 <KUR>  
A;Cross-references: UNIPROT:Q92G90; UNIPARC:UPI00001252DE; GB:AE006914; PIDN:AAL03771.1;  
C;Genetics:  
A;Gene: acnA  
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)  
C;Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 62.3%; Score 38; DB 2; Length 878;

Best Local Similarity 85.7%; Pred. No. 98;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7  
|||||  
Db 426 CTNTSNP 432

RESULT 15  
A71641  
aconitate hydratase (acna) RP799 - Rickettsia prowazekii  
C:Species: Rickettsia prowazekii  
C>Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 31-Dec-2004  
C:Accession: A71641  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sacheritz-Ponten, T.; Alsmark, U.  
Nature 396, 133-140, 1998  
A>Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499; PMID:9823893  
A:Accession: A71641  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-878 <AND>  
A:Cross-references: UNIPROT:Q9ZCF4; UNIPARC:UPI00001252DF; GB:AJ235269; NID  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: acna; RP799  
C:Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 878;  
Best Local Similarity 85.7%; Pred. No. 98;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7  
|||||  
Db 426 CTNTSNP 432

RESULT 16  
T27868  
hypothetical protein ZK455.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Dec-2004  
C:Accession: T27868  
R:White, S.  
submitted to the EMBL Data Library, November 1995  
A:Reference number: Z20432  
A:Accession: T27868  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-887 <WIL>  
A:Cross-references: UNIPROT:Q23500; UNIPARC:UPI00001252CD; EMBL:Z66567; PIDN:CAA91491.1;  
A:Experimental source: clone ZK455  
C:Genetics:  
A:Gene: CESP:ZK455.1  
A:Map position: X  
A:Introns: 31/1; 87/2; 133/2; 156/3; 426/2; 522/3; 715/3  
C:Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 887;  
Best Local Similarity 85.7%; Pred. No. 99;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7  
|||||  
Db 436 CTNTSNP 442

RESULT 17  
A44153  
aconitase hydratase (EC 4.2.1.3) - rabbit  
N:Alternate names: aconitase; ferritin mRNA repressor protein; iron-responsive element-b  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 31-Dec-2004

C:Accession: A44153; S47664; S47665  
R:Patino, M.M.; Walden, W.E.  
J. Biol. Chem. 267, 19011-19016, 1992  
A>Title: Cloning of a functional cDNA for the rabbit ferritin mRNA repressor protein. Den  
A:Reference number: A44153; MUID:92406828; PMID:1527028  
A:Accession: A44153  
A:Molecule type: mRNA; protein  
A:Residues: 1-889 <PAT>  
A:Cross-references: UNIPROT:Q01059; UNIPARC:UPI000012D87F; EMBL:M95815; NID:gl65029; PID  
A:Experimental source: liver  
A>Note: sequence extracted from NCBI backbone (NCBIN:113547, NCBIP:113546)  
R:Swenson, G.R.; Walden, W.E.  
Nucleic Acids Res. 22, 2627-2633, 1994  
A>Title: Localization of an RNA binding element of the iron responsive element binding pr  
A:Reference number: S47664; MUID:94316507; PMID:7518918  
A:Accession: S47664  
A:Molecule type: protein  
A:Residues: 133,'X',135-146 <SWE>  
A:Cross-references: UNIPARC:UPI00000172FE8  
A:Accession: S47665  
A:Molecule type: protein  
A:Residues: 624-635,'X',637-638 <SMW>  
A:Cross-references: UNIPARC:UPI0000172FE8  
C:Genetics:  
A:Gene: FRP  
C:Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)  
C:Keywords: 4Fe-4S; carbon-oxygen lyase; hydro-lyase; iron-sulfur protein; metalloprotein  
F:480-623/Domain: RNA binding #status predicted <RNA>  
F:125,126,178,205,207,302,778/Active site: Asp, His, His, Asp, His, Glu, Ser #status pred  
F:437,503,506/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 1; Length 889;  
Best Local Similarity 85.7%; Pred. No. 99;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7  
|||||  
Db 437 CTNTSNP 443

RESULT 18  
A44154  
aconitase hydratase (EC 4.2.1.3) - rat  
N:Alternate names: aconitase; iron-responsive element-binding protein  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 27-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C:Accession: A44154  
R:Yu, Y.; Radisky, E.; Leibold, E.A.  
J. Biol. Chem. 267, 19005-19010, 1992  
A>Title: The iron-responsive element binding protein. Purification, cloning, and regulati  
A:Reference number: A44154; MUID:92406827; PMID:1527027  
A:Accession: A44154  
A>Status: preliminary  
A:Molecule type: mRNA; protein  
A:Residues: 1-889 <YU1>  
A:Cross-references: UNIPARC:UPI00000176090  
A:Experimental source: liver  
A>Note: sequence extracted from NCBI backbone (NCBIN:113545, NCBIP:113544)  
C:Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)  
C:Keywords: 4Fe-4S; carbon-oxygen lyase; hydro-lyase; iron-sulfur protein; metalloprotein  
F:480-623/Domain: RNA binding #status predicted <RNA>  
F:125,126,178,205,207,302,778/Active site: Asp, His, His, Asp, His, Glu, Ser #status pred  
F:437,503,506/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 2; Length 889;  
Best Local Similarity 85.7%; Pred. No. 99;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7  
|||||  
Db 437 CTNTSNP 443

```

RESULT 19
S26403
aconitate hydratase (EC 4.2.1.13) - human
N;Alternate names: aconitase; iron regulatory factor; iron-responsive element-binding protein
C;Species: Homo sapiens (man)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 31-Dec-2004
C;Accession: S26403; A36203; S22147
R;Hirling, H.; Emery-Goodman, A.; Thompson, N.; Neupert, B.; Seiser, C.; Kuehn, L.C.
Nucleic Acids Res. 20, 33-39, 1992
A;Title: Expression of active iron regulatory factor from a full-length human cDNA by in
A;Reference number: S26403; MUID:92150156; PMID:1738601
A;Accession: S26403
A;Molecule type: mRNA
A;Residues: 1-889 <HIR>
A;Cross-references: UNIPROT:P21399; UNIPARC:UPI000012D87E; EMBL:Z11559; NID:g33962; PIDN
R;Rouault, T.A.; Tang, C.K.; Kaptain, S.; Burgees, W.H.; Haile, D.J.; Samaniego, F.; McB
Proc. Natl. Acad. Sci. U.S.A. 87, 7958-7962, 1990
A;Title: Cloning of the cDNA encoding an RNA regulatory protein--the human iron-respons
A;Reference number: A36203; MUID:91045916; PMID:2172968
A;Accession: A36203
A;Molecule type: mRNA
A;Residues: 100-531, 'II', 534, 'TGILKAELY', 544-711, 'PEVWTPSHGHEHLPTL', 728-889 <ROU>
A;Cross-references: UNIPARC:UPI000017608F; GB:M37836
C;Genetics:
A;Gene: GDB:IREB1
A;Cross-references: GDB:125344; OMIM:147581
A;Map position: 9pter-9qter
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)
C;Keywords: 4Fe-4S; carbon-oxygen lyase; hydro-lyase; iron-sulfur protein; metalloprotei
F;480-623/Domain: RNA binding #status predicted <RNA>
F;125,126,178,205,207,302,778/Active site: Asp, His, His, Asp, His, Glu, Ser #status pre
F;437,503,506/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 2; Length 889;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
|||||
Db 437 CTNTSNP 443

RESULT 20
S18720
aconitate hydratase (EC 4.2.1.13) - mouse
N;Alternate names: aconitase; iron-responsive element-binding protein
C;Species: Mus musculus (house mouse)
C;Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 31-Dec-2004
C;Accession: S18720
R;Philpott, C.C.; Rouault, T.A.; Klausner, R.D.
Nucleic Acids Res. 19, 6333, 1991
A;Title: Sequence and expression of the murine iron-responsive element binding protein.
A;Reference number: S18720; MUID:92066494; PMID:1956798
A;Accession: S18720
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-889 <PHI>
A;Cross-references: UNIPROT:P28271; UNIPARC:UPI000002805A; EMBL:X61147; NID:g52735; PIDN
A;Note: The nucleotide sequence was submitted to the EMBL Data Library, July 1991
C;Genetics:
A;Gene: irebP; acol
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)
C;Keywords: 4Fe-4S; carbon-oxygen lyase; glyoxylate bypass; hydro-lyase; iron-sulfur pro
F;480-623/Domain: RNA binding #status predicted <RNA>
F;125,126,178,205,207,302,778/Active site: Asp, His, His, Asp, His, Glu, Ser #status pre
F;437,503,506/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 2; Length 889;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
|||||
Db 437 CTNTSNP 443

RESULT 21
AH0270
aconitate hydratase (EC 4.2.1.13) [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 31-Dec-2004
C;Accession: AH0270
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
illo, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AH0270
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-890 <KUR>
A;Cross-references: UNIPROT:Q8ZEFL; UNIPARC:UPI00000DC909; GB:AL590842; PIDN:CAC91028.1;
C;Genetics:
A;Gene: acnA
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)
C;Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 62.3%; Score 38; DB 2; Length 890;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
|||||
Db 435 CTNTSNP 441

RESULT 22
AH054
aconitate hydratase 1 (citrate hydro-lyase 1) [imported] - Salmonella enterica subsp. ent
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: This species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 31-Dec-2004
C;Accession: AH0654
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AH0654
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-891 <PAR>
A;Cross-references: UNIPARC:UPI0000059F3C; GB:AL513382; PIDN:CAD08419.1; PID:g16502462;
C;Genetics:
A;Gene: STY1339
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 891;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7
|||||
Db 435 CTNTSNP 441

RESULT 23
B48642
aconitate hydratase (EC 4.2.1.13) - Legionella pneumophila
C;Species: Legionella pneumophila
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 31-Dec-2004
C;Accession: B48642
R;Mengaude, J.M.; Horwitz, M.A.

```

J. Bacteriol. 175, 5666-5676, 1993  
A;Title: The major iron-containing protein of *Legionella pneumophila* is an aconitase hom  
A;Reference number: A48642; MUID:93374864; PMID:8366052  
A;Accession: B48642  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-891 <MEN>  
A;Cross-references: UNIPROT:P37032; UNIPARC:UPI00001252DB; GB:L22081; NID:G348943; PIDN:  
A;Experimental source: strain Philadelphia 1, substrain pneumophila  
C;Genetics:  
A;Gene: acn  
C;Function:  
A;Description: reversibly catalyzes the hydration of cis-aconitate to citrate and also t  
A;Pathway: tricarboxylic acid cycle  
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)  
C;Keywords: 4Fe-4S; carbon-oxygen lyase; hydro-lyase; iron-sulfur protein; metalloprotei  
F;478-620/Domain: RNA binding #status predicted <RNA>  
F;130,131,183,214,216,311,776/Active site: Asp, His, His, Asp, His, Glu, Ser #status pre  
F;435,501,504/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 2; Length 891;  
Best Local Similarity 85.7%; Pred. No. 99;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7  
|||||  
Db 435 CTNTSNP 441

RESULT 24  
Aconitase hydratase 1 [imported] - *Escherichia coli* (strain O157:H7, substrain RIMD 050995  
C;Species: *Escherichia coli*  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 31-Dec-2004  
C;Accession: A90860  
R;Hayaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genc  
A;Reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: A90860  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-891 <HAY>  
A;Cross-references: UNIPROT:Q8X7C7; UNIPARC:UPI00000DOA85; GB:BA0000007; PIDN:BAH35272.1;  
A;Experimental source: strain O157:H7, substrain RIMD 0509952  
C;Genetics:  
A;Gene: Ecs1849  
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 891;  
Best Local Similarity 85.7%; Pred. No. 99;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7  
|||||  
Db 435 CTNTSNP 441

RESULT 25  
Aconitase hydratase 1 [imported] - *Escherichia coli* (strain O157:H7, substrain EDL933)  
C;Species: *Escherichia coli*  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Dec-2004  
C;Accession: B85759  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: B85759  
A;Status: preliminary  
A;Molecule type: DNA

A;Residues: 1-891 <STO>  
A;Cross-references: UNIPROT:Q8X7C7; UNIPARC:UPI00000DOA85; GB:AE005174; NID:G12515516; P  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: acnA  
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 891;  
Best Local Similarity 85.7%; Pred. No. 99;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7  
|||||  
Db 435 CTNTSNP 441

RESULT 26  
G64875  
Aconitase hydratase (EC 4.2.1.3) - *Escherichia coli* (strain K-12)  
C;Species: *Escherichia coli*  
C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 31-Dec-2004  
C;Accession: G64875; S22375; A49756  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Coi  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of *Escherichia coli* K-12.  
A;Reference number: A64720; MUID:97426617; PMID:9278503  
A;Accession: G64875  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-891 <BLAT>  
A;Cross-references: UNIPROT:P25516; UNIPARC:UPI00000480C9; GB:AE000225; GB:U00096; NID:G1  
A;Experimental source: strain K-12, substrain MG1655  
R;Prodromou, C.; Artymluk, P.J.; Guest, J.R.  
Eur. J. Biochem. 204, 599-609, 1992  
A;Title: The aconitase of *Escherichia coli*. Nucleotide sequence of the aconitase gene and  
isopropylmalate isomerase.  
A;Reference number: S22374; MUID:92174916; PMID:1541275  
A;Accession: S22375  
A;Molecule type: DNA  
A;Residues: 1-521, 'G', 523-891 <PRO>  
A;Cross-references: UNIPARC:UPI000016ED57; EMBL:X60293; NID:G40894; PIDN:CAA42834.1; PID:  
A;Experimental source: strain K-12, substrain W3110  
R;Prodromou, C.; Haynes, M.J.; Guest, J.R.  
J. Gen. Microbiol. 137, 2505-2515, 1991  
A;Title: The aconitase of *Escherichia coli*: purification of the enzyme and molecular clor  
A;Reference number: A49756; MUID:92148368; PMID:1838390  
A;Accession: A49756  
A;Molecule type: protein  
A;Residues: 'X', 3-15, 'X', 17, 'X', 19 <PR2>  
A;Cross-references: UNIPARC:UPI0000176094  
C;Genetics:  
A;Gene: acnA; acn  
C;Function:  
A;Description: reversibly catalyzes the hydration of cis-aconitate to citrate and also t  
A;Pathway: tricarboxylic acid cycle  
C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)  
C;Keywords: 4Fe-4S; carbon-oxygen lyase; hydro-lyase; iron-sulfur protein; metalloprotein  
F;2-891/Product: aconitate hydratase #status predicted <MAI>  
F;478-620/Domain: RNA binding #status predicted <RNA>  
F;128,129,181,212,214,309,776/Active site: Asp, His, His, Asp, His, Glu, Ser #status pred  
F;435,501,504/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 2; Length 891;  
Best Local Similarity 85.7%; Pred. No. 99;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7  
|||||  
Db 435 CTNTSNP 441

RESULT 27

AI3483  
 A;Title: aconitate hydratase (EC 4.2.1.3) [imported] - Brucella melitensis (strain 16M)  
 C;Species: Brucella melitensis  
 C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 31-Dec-2004  
 C;Accession: AI3483  
 R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova, A.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
 A;Reference number: AD3252; PMID:11756688  
 A;Accession: AI3483  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-895 <KUR>  
 A;Cross-references: UNIPROT:Q8YEM3; UNIPARC:UPI0000058222; GB:AE008917; PIDN:AAL53036.1;  
 A;Experimental source: strain 16M  
 C;Genetics:  
 A;Gene: BME11855  
 A;Map position: I  
 C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)  
 C;Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 62.3%; Score 38; DB 2; Length 895;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7  
 |||||  
 Db 436 CTNTSNP 442

RESULT 28  
 AD2906  
 A;Title: aconitate hydratase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 C;Species: Agrobacterium tumefaciens  
 C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 31-Dec-2004  
 C;Accession: AD2906  
 R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McCllellan, Science 294, 2317-2323, 2001  
 A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.W.  
 A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A;Reference number: AB2577; MUID:21608550; PMID:11743193  
 A;Accession: AD2906  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-897 <KUR>  
 A;Cross-references: UNIPROT:Q8UC13; UNIPARC:UPI00000D1FB5; GB:AE008688; PIDN:AAL43666.1;  
 A;Experimental source: strain C58 (Dupont)  
 C;Genetics:  
 A;Gene: acna  
 A;Map position: circular chromosome  
 C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 897;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7  
 |||||  
 Db 437 CTNTSNP 443

RESULT 29  
 E97681  
 A;Title: aconitate hydratase (citrate hydro-lyase) (aconitase) [imported] - Agrobacterium tumefaciens  
 C;Species: Agrobacterium tumefaciens  
 C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 31-Dec-2004  
 C;Accession: E97681  
 R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001  
 A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
 A;Reference number: A97359; MUID:21608551; PMID:11743194  
 A;Accession: E97681  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-897 <KUR>  
 A;Cross-references: UNIPROT:Q8UC13; UNIPARC:UPI00000D1FB5; GB:AE007869; PIDN:AAK88406.1;  
 C;Genetics:  
 A;Gene: AGR C 4866  
 A;Map position: circular chromosome  
 C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 897;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7  
 |||||  
 Db 437 CTNTSNP 443

RESULT 30  
 B84471  
 A;Title: cytoplasmic aconitate hydratase [imported] - Arabidopsis thaliana  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 31-Dec-2004  
 C;Accession: B84471  
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Neebawong, M.; Koo, H.; Moffat, K.S.; Cronlin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999  
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A;Reference number: A84420; MUID:20083487; PMID:10617197  
 A;Accession: B84471  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-898 <STO>  
 A;Cross-references: UNIPROT:Q9SIB9; UNIPARC:UPI000009F65C; GB:AE002093; NID:g4586021; PII C;Genetics:  
 A;Gene: At2g05710  
 A;Map position: 2  
 C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 898;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7  
 |||||  
 Db 441 CTNTSNP 447

RESULT 31  
 T04693  
 A;Title: aconitate hydratase (EC 4.2.1.3), cytosolic - Arabidopsis thaliana  
 N;Alternate names: aconitase; protein F4B14.100  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 31-Dec-2004  
 C;Accession: T04693; S49847  
 R;Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.X. submitted to the Protein Sequence Database, October 1998  
 A;Reference number: 215380  
 A;Accession: T04693  
 A;Molecule type: DNA  
 A;Residues: 1-898 <BEV>  
 A;Cross-references: UNIPROT:Q42560; UNIPARC:UPI00001252CC; EMBL:AL031986  
 A;Experimental source: cultivar Columbia; BAC clone F4B14  
 R;Peyret, P.; Persz, P.; Alric, M. submitted to the EMBL Data Library, November 1994  
 A;Description: Plant aconitase enzyme.  
 A;Reference number: S49847  
 A;Accession: S49847

A:Molecule type: mRNA

A:Residues: 'NFLSSENRSNLYFASSLDLYLSS', 1-129; 'Y', 131-231; 'RP', 234-714; 'VAVVMRLREH', 728-848

A:Cross-references: UNIPARC:UPI000011D718; EMBL:X82839; NID:G599624; PID:G599625

A:Experimental source: variety ecotype Columbia; unripened pods

A:Note: the extension at the 5' end results from translation of the 5' untranslated region

A:Note: the differences near the carboxyl end are due to a frameshift error

C:Genetics:

A:Gene: ACO

A:Map position: 4

A:Introns: 2/1; 37/1; 90/3; 121/3; 183/3; 237/3; 286/1; 333/3; 352/3; 379/2; 404/3; 466/3

A:Note: F4B14.100

C:Function:

A:Description: reversibly catalyzes the hydration of cis-aconitate to citrate and also citrate to cis-aconitate

C:Pathway: glyoxylate bypass

C:Superfamily: aconitate A/homoaconitase (aconitate hydratase 1)

C:Keywords: 4Fe-4S; carbon-oxygen lyase; glyoxylate bypass; hydro-lyase; iron-sulfur protein

F:484-627/Domains: RNA binding #status predicted <RNA>

F:441,507,510/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 2; Length 898;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7  
|||||

Db 441 CTNTSNP 447

RESULT 32

T10101

aconitate hydratase (EC 4.2.1.3) - cucurbit

N:Alternate names: aconitase

C:Species: Cucurbita sp. (cucurbit)

C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 31-Dec-2004

C:Accession: T10101

R:Hayashi, M.; De Bellis, L.; Alpi, A.; Nishimura, M.  
Plant Cell Physiol. 36, 669-680, 1995

A:Title: Cytosolic aconitase participates in the glyoxylate cycle in etiolated pumpkin o

A:Reference number: Z16948; MUID:95368272; PMID:7640891

A:Accession: T10101

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-898 <HAY>

A:Cross-references: UNIPARC:UPI00001252CE; EMBL:D29629; NID:G868002; PIDN:BAA06108.1; PI

A:Experimental source: cv. Kurokawa Amakuri Nankin; etiolated cotyledons

C:Function:

A:Description: catalyzes the reversible hydration of cis-aconitate to citrate and also cit

A:Pathway: tricarboxylic acid cycle

C:Superfamily: aconitate A/homoaconitase (aconitate hydratase 1)

C:Keywords: carbon-oxygen lyase; cytosol; hydro-lyase

Query Match 62.3%; Score 38; DB 2; Length 898;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7  
|||||

Db 441 CTNTSNP 447

RESULT 33

AI1279

aconitate hydratases homolog citB [imported] - Listeria monocytogenes (strain EGD-e)

C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 31-Dec-2004

C:Accession: AI1279

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001

A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Smoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlard,  
A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: A11279  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-900 <GLA>  
A:Cross-references: UNIPROT:Q8Y6P3; UNIPARC:UPI0000055009; GB:NC\_003210; PIDN:CAC99719.1  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: citB  
C:Superfamily: aconitase A/homoaconitase (aconitase hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 900;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7  
|||||  
Db 443 CTNTSNP 449

RESULT 34  
A11642  
aconitase hydratases homolog citB [imported] - *Listeria innocua* (strain Clip11262)  
C:Species: *Listeria innocua*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 31-Dec-2004  
C:Accession: A11642  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H.; D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maltournam, A.; Matok, C.; Schluter, J.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: A11642  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-900 <GLA>  
A:Cross-references: UNIPROT:Q92B68; UNIPARC:UPI00000CC61F; GB:AL592022; PIDN:CAC96913.1  
A:Experimental source: strain Clip11262  
C:Genetics:  
A:Gene: citB  
C:Superfamily: aconitase A/homoaconitase (aconitase hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 900;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7  
|||||  
Db 443 CTNTSNP 449

RESULT 35  
F89910  
aconitase hydratase [imported] - *Staphylococcus aureus* (strain N315)  
C:Species: *Staphylococcus aureus*  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 31-Dec-2004  
C:Accession: F89910  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Oi, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shibata, T.; Hattori, M.; Ogasawara, N.; Hayaashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
A:Reference number: AB9758; MUID:21311952; PMID:11418146  
A:Accession: F89910  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-901 <KUR>  
A:Cross-references: UNIPROT:Q99UC8; UNIPARC:UPI00000549F4; GB:BA000018; PID:G13701147; P3  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: citB

C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 901;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7  
 ||||| ||

Db 443 CTNTSNP 449

# RESULT 36

A87704  
 aconitate hydratase 1 [imported] - Caulobacter crescentus  
 C;Species: Caulobacter crescentus  
 C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 31-Dec-2004  
 C;Accession: A87704  
 R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A;Title: Complete Genome Sequence of Caulobacter crescentus.  
 A;Reference number: A87249; MUID:21173698; PMID:11259647  
 A;Accession: A87704  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-903 <STO>  
 A;Cross-references: UNIPROT:Q9A299; UNIPARC:UPI00000C7B68; GB:AE005673; NID:gl13425425; F  
 C;Genetics:  
 A;Gene: CC3667  
 C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 903;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7  
 ||||| ||

Db 444 CTNTSNP 450

# RESULT 37

G75362  
 aconitate hydratase - Deinococcus radiodurans (strain R1)  
 C;Species: Deinococcus radiodurans  
 C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Dec-2004  
 C;Accession: G75362  
 R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A;Reference number: A75250; MUID:20036896; PMID:10567266  
 A;Accession: G75362  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-906 <WHI>  
 A;Cross-references: UNIPROT:Q9RTN7; UNIPARC:UPI00000D3EC3; GB:AE002013; GB:AE0000513; NID  
 A;Experimental source: strain R1  
 C;Genetics:  
 A;Gene: DR1720  
 A;Map position: 1  
 C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 906;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7  
 ||||| ||

Db 441 CTNTSNP 447

# RESULT 38

T04820  
 aconitate hydratase (EC 4.2.1.3) F10M23.310 - Arabidopsis thaliana  
 N;Alternate names: protein F10M23.310  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 31-Dec-2004  
 C;Accession: T04820  
 R;Bevan, M.; Lechamny, A.; Chedford, F.; Krivitzky, M.; Kreis, M.; Hoheisel, J.; Mewes, H.  
 submitted to the Protein Sequence Database, February 1999  
 A;Reference number: Z15385  
 A;Accession: T04820  
 A;Molecule type: DNA  
 A;Residues: 1-907 <BEV>  
 A;Cross-references: UNIPROT:Q9S236; UNIPARC:UPI00000AA6C2; EMBL:AL035440  
 A;Experimental source: cultivar Columbia; BAC clone F10M23  
 C;Genetics:  
 A;Map position: 4  
 A;Introns: 46/1; 99/3; 130/3; 192/3; 246/3; 295/1; 363/3; 388/2; 413/3; 475/3; 498/2; 531  
 A;Note: F10M23.310  
 C;Function:  
 A;Description: reversibly catalyzes the hydration of cis-aconitate to citrate and also to  
 A;Pathway: glyoxylate bypass  
 C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)  
 C;Keywords: 4Fe-4S; carbon-oxygen lyase; glyoxylate bypass; hydro-lyase; iron-sulfur prot

Query Match 62.3%; Score 38; DB 2; Length 907;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7  
 ||||| ||

Db 450 CTNTSNP 456

# RESULT 39

G82824  
 aconitase XF0290 [imported] - Xylella fastidiosa (strain 945c)  
 C;Species: Xylella fastidiosa  
 C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 31-Dec-2004  
 C;Accession: G82824  
 R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
 Nature 406, 151-157, 2000  
 A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A;Reference number: A82515; MUID:20365717; PMID:10910347  
 A;Note: for a complete list of authors see reference number A59328 below  
 A;Accession: G82824  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-908 <SIM>  
 A;Cross-references: UNIPROT:Q9PGK8; UNIPARC:UPI00000C236F; GB:AE003883; GB:AE003849; NID:  
 A;Experimental source: strain 945c  
 R;Simpson, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H.  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigre  
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.  
 A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 , F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira  
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A;Reference number: A59328  
 A;Contents: annotation  
 C;Genetics:  
 A;Gene: XF0290  
 C;Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 908;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;



QY 1 CTNTDNP 7  
 |||||  
 Db 443 CTNTSNP 449

RESULT 40  
 G69599  
 aconitate hydratase (EC 4.2.1.3) citB - Bacillus subtilis  
 N/Alternate names: aconitase  
 C/Species: Bacillus subtilis  
 C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 31-Dec-2004  
 C/Accession: G69599; A27085  
 R/Kunst, S.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
 A.; Ehrlich, S.D.; Emerson, F.F.; Encian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle  
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
 A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
 A/Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Ser  
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
 A/Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
 A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A/Reference number: A69580; MUID:98044033; PMID:9384377  
 A/Accession: G69599  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-909 <KUN>  
 A/Cross-references: UNIPROT:P09339; UNIPARC:UPI0000060488; GB:Z99113; GB:AL009126; NID:9  
 A/Experimental source: strain 168  
 R/Dingman, D.W.; Sonenshein, A.L.  
 J. Bacteriol. 169, 3062-3067, 1987  
 A/Title: Purification of aconitase from Bacillus subtilis and correlation of its N-termi  
 A/Reference number: A27085; MUID:87250270; PMID:3110133  
 A/Accession: A27085  
 A/Molecule type: DNA  
 A/Residues: 1-28 'I', 30-42, 877-879 <DIN>  
 A/Cross-references: UNIPARC:UPI0000176093  
 A/Note: the authors translated the codon ATC for residue 29 as Tyr  
 C/Genetics:  
 A/Gene: citB  
 C/Function:  
 A/Description: reversibly catalyzes the hydration of cis-aconitate to citrate and also t  
 A/Pathway: tricarboxylic acid cycle  
 C/Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)  
 C/Keywords: 4Fe-4S; carbon-oxygen lyase; hydro-lyase; iron-sulfur protein; metalloprotei  
 F:493-636/Domain: RNA binding #status predicted <RNA>  
 F:135.136.188.219.221.316.792/Active site: Asp, His, His, Asp, His, Glu, Ser #status pre  
 F:450.516.519/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 2; Length 909;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7  
 |||||  
 Db 450 CTNTSNP 456

RESULT 41  
 B83451  
 aconitate hydratase 1 PA1562 [imported] - Pseudomonas aeruginosa (strain PAO1)  
 C/Species: Pseudomonas aeruginosa  
 C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2004  
 C/Accession: B83451  
 R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000

A/Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen  
 A/Reference number: A82950; MUID:20437337; PMID:10984043  
 A/Accession: B83451  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-910 <STO>  
 A/Cross-references: UNIPROT:O913F5; UNIPARC:UPI00000C53E0; GB:AE004584; GB:AE004091; NID:  
 A/Experimental source: strain PAO1  
 C/Genetics:  
 A/Gene: acnA; PA1562  
 C/Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 910;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7  
 |||||  
 Db 454 CTNTSNP 460

RESULT 42  
 F70873  
 aconitate hydratase (EC 4.2.1.3) - Mycobacterium tuberculosis (strain H37RV)  
 C/Species: Mycobacterium tuberculosis  
 C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 31-Dec-2004  
 C/Accession: F70873  
 R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
 i Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A/Reference number: A70500; MUID:98295987; PMID:9634230  
 A/Accession: F70873  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-943 <COL>  
 A/Cross-references: UNIPROT:O53166; UNIPARC:UPI00000D0F85; GB:AL021184; GB:AL123456; NID:  
 A/Experimental source: strain H37RV  
 C/Genetics:  
 A/Gene: acn  
 C/Function:  
 A/Description: reversibly catalyzes the hydration of cis-aconitate to citrate and also t  
 A/Pathway: tricarboxylic acid cycle  
 C/Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)  
 C/Keywords: 4Fe-4S; carbon-oxygen lyase; hydro-lyase; iron-sulfur protein; metalloprotei  
 F:522-665/Domain: RNA binding #status predicted <RNA>  
 F:126.127.179.206.208.303.823/Active site: Asp, His, His, Asp, His, Glu, Ser #status pre  
 F:479.545.548/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 62.3%; Score 38; DB 2; Length 943;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 7  
 |||||  
 Db 479 CTNTSNP 485

RESULT 43  
 G87135  
 aconitate hydratase [imported] - Mycobacterium leprae  
 C/Species: Mycobacterium leprae  
 C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 31-Dec-2004  
 C/Accession: G87135  
 R/Cole, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho  
 R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
 sam, M.A.; Rutherford, K.M.  
 Nature 409, 1007-1011, 2001  
 A/Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq  
 A/Title: Massive gene decay in the leprosy bacillus.  
 A/Reference number: A86909; MUID:21128732; PMID:11234002

A:Accession: G87135  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-944 <STO>  
A:Cross-references: UNIPROT:Q9CBL3; UNIPARC:UPI00000C6DFD; GB:AL450380; NID:gl3093525; F  
C:Genetics:  
A:Gene: acn  
C:Superfamily: aconitase A/homoaconitase (aconitate hydratase 1)

Query Match 62.3%; Score 38; DB 2; Length 944;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 7  
||| |||  
Db 480 CINTSNP 486

RESULT 44  
A12205  
hypothetical protein alr3200 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C:Accession: A12205  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: A12205  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-201 <KUR>  
A:Cross-references: UNIPROT:Q8Y91; UNIPARC:UPI00000C36DD; GB:BA000019; PIDN:BA074899.1;  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr3200

Query Match 60.7%; Score 37; DB 2; Length 201;  
Best Local Similarity 60.0%; Pred. No. 36;  
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CTNTDNPYK 10  
| |||||  
Db 16 CQYRDNPYE 25

RESULT 45  
A96670  
hypothetical protein F13011.1 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: A96670  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: A96670  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-323 <STO>  
A:Cross-references: UNIPROT:Q9XIS2; UNIPARC:UPI00000A5918; GB:AE005173; NID:gs042406; PI  
C:Genetics:  
A:Gene: F13011.1  
A:Map position: 1

Query Match 60.7%; Score 37; DB 2; Length 323;  
Best Local Similarity 66.7%; Pred. No. 56;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTNTDNPY 9  
|||||  
Db 10 CTNTQNRW 18

RESULT 46  
I52603  
MPS1 protein - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 09-Jul-2004  
C:Accession: I52603  
R:Spilebury, K.; O'Mara, M.A.; Wu, W.M.; Rowe, P.B.; Symonds, G.; Takayama, Y.  
Blood 85, 1620-1629, 1995  
A:Title: Isolation of a novel macrophage-specific gene by differential cDNA analysis.  
A:Reference number: I52603; MUID:95195232; PMID:7888681  
A:Accession: I52603  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-661 <RES>  
A:Cross-references: UNIPROT:Q61889; UNIPARC:UPI000002910D; GB:IL20315; NID:g431419; PIDN:1

Query Match 60.7%; Score 37; DB 2; Length 661;  
Best Local Similarity 55.6%; Pred. No. 1.1e+02;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTNTDNPY 9  
||| |||  
Db 349 CTNVDSNPF 357

RESULT 47  
F64112  
malate dehydrogenase (oxaloacetate-decarboxylating) homolog - Haemophilus influenzae (str  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 05-Oct-2004  
C:Accession: F64112  
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, C.  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: F64112  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-756 <TIGR>  
A:Cross-references: UNIPROT:P43837; UNIPARC:UPI00000512CB; GB:U32804; GB:IL42023; NID:gl5;  
C:Superfamily: malic enzyme with phosphate acetyl/butaryl transferase domain

Query Match 60.7%; Score 37; DB 2; Length 756;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 NTDNPRYK 10  
|||||  
Db 501 NEDNPRYE 508

RESULT 48  
E88412  
protein C44F1.5 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: E88412  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology

A;Reference number: A75000; MUID:99069613; PMID:9851916  
A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_elegans/  
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A;Accession: E88412  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-805 <S>O>  
A;Cross-references: UNIPROT:Q18628; UNIPARC:UPI00000769AA; GB:chr\_III; PIDN:CAA88885.1;  
C;Genetics:  
A;Gene: C44F1.5  
A;Map position: 3

Query Match 60.7%; Score 37; DB 2; Length 805;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NTQDPRYK 10  
| | | | |  
| | | | |  
Db 694 NTQDPRYE 701

## RESULT 49

T19936  
hypothetical protein C44F1.5 - Caenorhabditis elegans (fragment)  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T19936; T25216  
R;Percy, C.  
submitted to the EMBL Data Library, April 1995  
A;Reference number: Z19199  
A;Accession: T19936  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-805 <W>I>  
A;Cross-references: UNIPROT:Q18628; UNIPARC:UPI00000769AA; EMBL:Z49067; PIDN:CAA88853.1;  
A;Experimental source: clone C44F1  
R;Percy, C.  
submitted to the EMBL Data Library, April 1995  
A;Reference number: Z19997  
A;Accession: T25216  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-805 <W>I2>  
A;Cross-references: UNIPARC:UPI00000769AA; EMBL:Z49072; PIDN:CAA88885.1; GSPDB:GN000021;  
A;Experimental source: clone T24A11  
C;Genetics:  
A;Gene: CESP:C44F1.5  
A;Map position: 3  
A;Introns: 10/3; 60/3; 94/3; 227/2; 271/2; 314/2; 402/3; 434/3; 682/1

Query Match 60.7%; Score 37; DB 2; Length 805;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NTQDPRYK 10  
| | | | |  
| | | | |  
Db 694 NTQDPRYE 701

## RESULT 50

T37523  
probable oxoprolinase - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C;Accession: T37523  
R;Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, December 1995  
A;Reference number: Z21720  
A;Accession: T37523  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1260 <CON>  
A;Cross-references: UNIPROT:Q10093; UNIPARC:UPI000013A13B; EMBL:Z68166; PIDN:CAA92315.1;

A;Experimental source: strain 972h-; cosmid c11D3  
C;Genetics:  
A;Gene: SPDB:SPAC11D3.14c  
A;Map position: 1  
C;Superfamily: hypothetical protein YKL215c

Query Match 60.7%; Score 37; DB 2; Length 1260;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTNTDNPY 9  
| | | | |  
| | | | |  
Db 1220 CSNPDPFY 1228

Search completed: June 5, 2006, 12:54:04  
Job time : 21.0959 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:32:17 ; Search time 87.6712 Seconds  
(without alignments)  
105.510 Million cell updates/sec

Title: US-10-645-659A-6

Perfect score: 61

Sequence: 1 CTNTDNPYK 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : Uniprot\_7.2.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	100.0	543	1	HPSE HUMAN
2	49	80.3	545	1	Q9y251 homo sapien
3	43	70.5	567	2	Q3R2P1_XYLFA
4	43	70.5	608	2	Q3R890_XYLFA
5	43	70.5	861	2	Q3QL63_9GAMM
6	43	70.5	862	2	Q33T54_9GAMM
7	43	70.5	862	2	Q4ZUP2_PSEU2
8	43	70.5	862	2	Q88KF4_PSEPK
9	43	70.5	863	2	Q2X9A9_PSEPU
10	43	70.5	863	2	Q48JZ8_PSE14
11	43	70.5	863	2	Q2P7A3_XANOR
12	43	70.5	863	2	Q87P72_VIBPA
13	43	70.5	863	2	Q883R1_PSESM
14	43	70.5	863	2	Q8PND4_XANAC
15	43	70.5	864	2	Q2T374_BURTH
16	43	70.5	864	2	Q3BWH4_XANC5
17	43	70.5	864	2	Q454U1_9BURK
18	43	70.5	864	2	Q4LQNS_9BURV
19	43	70.5	864	2	Q41GN5_BURKH
20	43	70.5	864	2	Q4URR6_XANC8
21	43	70.5	864	2	Q8VPS8_9BURK
22	43	70.5	864	2	Q39BA1_BURS3
23	43	70.5	864	2	Q3JH08_BURP1
24	43	70.5	864	2	Q3KFE7_PSEPF
25	43	70.5	864	2	Q4KFK1_PSEF5
26	43	70.5	864	2	Q5QZU0_IDILO
27	43	70.5	864	2	Q62A69_EURMA
28	43	70.5	864	2	Q63NU1_BURPS
29	43	70.5	864	2	Q8PBT6_XANCP
30	43	70.5	865	2	Q3F693_9BURK
31	43	70.5	865	2	Q3RQY1_RALME

32	43	70.5	865	2	Q8XTI8_RALSO
33	43	70.5	867	2	Q360W0_shewanella
34	43	70.5	867	2	Q36CY9_shewanella
35	43	70.5	867	2	Q3P203_9GAMM
36	43	70.5	867	2	Q4J5X3_AZOV1
37	43	70.5	867	2	Q8EJW3_shewanella
38	43	70.5	868	2	Q3CT85_pseudoalter
39	43	70.5	868	2	Q480I7_colwellia p
40	43	70.5	868	2	Q7NWD6_chromobacte
41	43	70.5	868	2	Q915E4_pseudomonas
42	43	70.5	868	2	Q9JT05_NEIMA
43	43	70.5	868	2	Q9K0X3_NEIMB
44	43	70.5	868	2	Q9KSC0_VIBCH
45	43	70.5	869	2	Q3IH91_PSEHT
46	43	70.5	869	2	Q937N8_RALEU
47	43	70.5	869	2	Q2KW75_BORAV
48	43	70.5	870	2	Q3NW39_SHEFR
49	43	70.5	871	2	Q2WXY8_9GAMM
50	43	70.5	871	2	Q2ZVF7_SHEPU
51	43	70.5	871	2	Q3QAM9_9GAMM
52	43	70.5	871	2	Q7ML96_VIBVU
53	43	70.5	871	2	Q8D975_VIBVU
54	43	70.5	872	2	Q34VQ9_9GAMM
55	43	70.5	874	2	Q36Q22_MARHY
56	43	70.5	877	2	Q8F8W5_AC1AD
57	43	70.5	881	2	Q470K8_RALEJ
58	43	70.5	884	2	Q2ZAX7_9GAMM
59	43	70.5	885	2	Q3GR04_9GAMM
60	43	70.5	885	2	Q4FSP5_PSYAR
61	43	70.5	897	2	Q7VM95_BORPE
62	43	70.5	897	2	Q7WSQ5_BORPA
63	43	70.5	897	2	Q7WD91_BORBR
64	43	70.5	901	2	Q3RGR7_XYLFA
65	43	70.5	913	2	Q5H4H3_XANOR
66	43	70.5	939	2	Q7SD71_NEUCR
67	42	68.9	791	2	Q61KT7_CABER
68	41	67.2	82	2	Q66LT3_BPT5
69	41	67.2	87	2	Q6QGF5_BPT5
70	41	67.2	469	2	Q6FUB4_CANGA
71	41	67.2	558	2	Q333X5_SPJJD
72	41	67.2	574	2	Q333X6_SPJJD
73	41	67.2	574	2	Q333X7_9RODE
74	41	67.2	574	2	Q333X8_9RODE
75	41	67.2	574	2	Q333X9_9RODE
76	40	65.6	183	2	Q3CY93_STRAG
77	40	65.6	184	2	Q8XT50_RALSO
78	40	65.6	297	1	DAPA_LACLA
79	40	65.6	297	2	Q54A52_LACLA
80	40	65.6	338	2	Q2XSC0_MYTED
81	40	65.6	490	2	Q3DDM5_STRAG
82	40	65.6	551	2	Q3DC63_STRAG
83	40	65.6	881	2	Q3D064_STRAG
84	40	65.6	881	2	Q3D9R8_STRAG
85	40	65.6	881	2	Q3DJX7_STRAG
86	40	65.6	881	2	Q3DRB7_STRAG
87	40	65.6	881	2	Q3JZL9_STR1
88	40	65.6	881	2	Q8DY23_STR45
89	40	65.6	881	2	Q8E3P2_STR43
90	40	65.6	960	1	SVV_BORBR
91	40	65.6	960	1	SVV_BORPA
92	40	65.6	960	1	SVV_BORPE
93	39	63.9	34	2	Q4T497_TETNG
94	39	63.9	145	2	Q6H520_ORYSA
95	39	63.9	235	2	Q4XEN9_PLACH
96	39	63.9	246	2	Q4X8D9_PLACH
97	39	63.9	259	2	Q97FY6_CLOAB
98	39	63.9	299	2	Q5FHS3_LACAC
99	39	63.9	536	1	HPSE_RAT
100	39	63.9	576	1	GLT12_MOUSE
101	39	63.9	576	2	Q60GT3_MOUSE
102	39	63.9	895	2	Q9Y040_PACLE
103	39	63.9	927	2	Q3BWM4_9ACTO
104	39	63.9	927	2	Q2JDF3_9ACTO

Q8xti8	ralstonia s
Q360w0	shewanella
Q36cy9	shewanella
Q3p203	shewanella
Q4j5x3	azotobacter
Q8ejw3	shewanella
Q3ct85	pseudoalter
Q480i7	colwellia p
Q7nwd6	chromobacte
Q915e4	pseudomonas
Q9jt05	neisseria m
Q9k0x3	neisseria m
Q9ksc0	vibrio chol
Q3ih91	pseudoalter
Q937n8	ralstonia e
Q2kw75	bordetella
Q3nw39	shewanella
Q2wxy8	shewanella
Q2zvf7	shewanella
Q3qam9	shewanella
Q7ml96	vibrio vuln
Q8d975	vibrio vuln
Q34vg9	alkalilimni
Q36q22	marinobacte
Q6f8w5	acinetobact
Q470k8	ralstonia e
Q2zax7	shewanella
Q3gr04	psychrobact
Q4fsp5	psychrobact
Q7vm95	bordetella
Q7wsq5	bordetella
Q7wd91	bordetella
Q3rgr7	xyella fas
Q5h4h3	xanthomonas
Q7sd71	neurospora
Q61kt7	caenorhabdi
Q66lt3	bacterioph
Q6qgf5	bacterioph
Q6fub4	candida gla
Q333x5	spalax juda
Q333x6	spalax juda
Q333x7	spalax carm
Q333x8	spalax gola
Q333x9	spalax gali
Q3cy93	streptococc
Q8xt50	ralstonia s
Q9cf61	lactococcus
Q54a52	lactococcus
Q2xsc0	mytilus edu
Q3ddm5	streptococc
Q3dc63	streptococc
Q3d064	streptococc
Q3d9r8	streptococc
Q3djx7	streptococc
Q3drb7	streptococc
Q3jzl9	streptococc
Q8dy23	streptococc
Q8e3p2	streptococc
Q7wm6	bordetella
Q7w6p3	bordetella
Q7w6k6	bordetella
Q4t497	tetraodon n
Q6h520	oryza sativ
Q4xen9	plasmodium
Q4x8d9	plasmodium
Q97fy6	clostridium
Q5fhs3	lactobacill
Q71rpl	rattus norv
Q8bt9	m polypepti
Q60gt3	m ucp-n-ace
Q9y040	paciastacu
Q3bwm4	frankia sp.
Q2jdf3	frankia sp.

105	39	63..9	938	2	Q4C1W5_CLOTH	Q4ciw5	clostridium	178	38	62..3	878	1	ACON_RICPR	Q9zcf4	rickettsia
106	39	63..9	938	2	Q5TIQ4_CLOTH	O5tiq4	clostridium	179	38	62..3	878	2	Q7P9X6_RICSI	Q7p9x6	rickettsia
107	39	63..9	2931	2	Q9W2C6_DROME	Q9w2c6	drosophila	180	38	62..3	878	2	Q4UK20_RICFE	Q4uk20	rickettsia
108	39	63..9	2968	2	Q8MLU9_DROME	O8mlu9	drosophila	181	38	62..3	878	2	Q6WV0_RICTE	Q6wv0	rickettsia
109	38	62..3	82	2	Q6V8S5_MALDO	O6v8s5	malus domes	182	38	62..3	879	2	Q70WQ0_THERTE	Q70wq0	thermoprote
110	38	62..3	255	2	Q7P631_FUSNV	Q7p631	fusobacteri	183	38	62..3	880	2	Q40LA7_DESAC	Q40la7	desulfurotom
111	38	62..3	256	2	Q444Z5_DROME	O444z5	drosophila	184	38	62..3	881	2	Q97CB3_THEVO	Q97cb3	thermoplasm
112	38	62..3	256	2	Q9VY18_DROME	Q9vy18	drosophila	185	38	62..3	881	2	Q5FE16_EHRRW	Q5fe16	ehrlichia r
113	38	62..3	261	2	Q3FW77_9BURK	Q3fw77	rhodopherax	186	38	62..3	881	2	Q5FGE6_EHRRG	Q5fge6	ehrlichia r
114	38	62..3	264	2	Q24234_DROME	Q24234	drosophila	187	38	62..3	883	2	Q3YQZ6_EHRCJ	Q3yqz6	ehrlichia c
115	38	62..3	278	2	Q3D7T9_STRAG	Q3d7t9	streptococc	188	38	62..3	884	2	Q3RWM6_EALME	Q3rwm6	raistoniam
116	38	62..3	308	2	Q9HHZ9_HALSA	Q9hhz9	halobacteri	189	38	62..3	887	1	ACOC_CAEEL	Q23500	caenorhabdi
117	38	62..3	317	2	Q6NB66_9PROT	O6neb6	magnetospir	190	38	62..3	887	2	Q10Q05_CAEHR	Q610q5	caenorhabdi
118	38	62..3	374	2	Q4W894_ELAQU	Q4w894	elaphu quad	191	38	62..3	887	2	Q5L290_STRT1	Q51290	streptococc
119	38	62..3	374	2	Q4W895_TRISI	Q4w895	trionyx sin	192	38	62..3	887	2	Q5M3U3_STRT2	Q5m3u3	streptococc
120	38	62..3	384	2	Q3D211_STRAG	Q3d211	streptococc	193	38	62..3	888	1	ACON_STRMU	Q59938	streptococc
121	38	62..3	384	2	Q3DCT0_STRAG	Q3dct0	streptococc	194	38	62..3	888	2	Q4ATR5_9BURK	Q4atr5	polaromonas
122	38	62..3	384	2	Q3DP88_STRAG	Q3dp88	streptococc	195	38	62..3	888	2	Q6ABV4_PROAC	Q6abv4	propionibac
123	38	62..3	384	2	Q3K0U0_STRA1	Q3k0u0	streptococc	196	38	62..3	889	1	IREB1_CHICK	Q90875	gallus gal
124	38	62..3	384	2	Q67A28_STRAG	Q67a28	streptococc	197	38	62..3	889	1	IREB1_HUMAN	P21399	homo sapien
125	38	62..3	384	2	Q67AX6_STRAG	O67ax6	streptococc	198	38	62..3	889	1	IREB1_MOUSE	P28271	mus musculu
126	38	62..3	384	2	Q93T11_STRAG	Q93t11	streptococc	199	38	62..3	889	1	IREB1_RABIT	O10159	oryctolagus
127	38	62..3	384	2	Q9AFH0_STRAG	Q9afh0	streptococc	200	38	62..3	889	1	IREB1_RAT	Q63270	rattus norv
128	38	62..3	384	2	Q9ALW5_STRAG	Q9alw5	streptococc	201	38	62..3	889	2	Q5VZA7_HUMAN	Q35vza7	homo sapien
129	38	62..3	384	2	Q9RPB9_STRAG	Q9rpb9	streptococc	202	38	62..3	889	2	Q5R511_PONPY	Q5r511	pongo pygma
130	38	62..3	384	2	Q9S0S6_STRAG	Q9s0s6	streptococc	203	38	62..3	889	2	Q300P3_STRSV	Q300p3	streptococc
131	38	62..3	384	2	Q8D2E4_STRA5	O8d2e4	streptococc	204	38	62..3	889	2	Q4JFF1_STRVR	Q4jff1	streptomyce
132	38	62..3	384	2	Q8E502_STRA3	O8e502	streptococc	205	38	62..3	889	2	Q4FL10_PELUB	O4fl10	pelagibacte
133	38	62..3	390	2	Q6BG80_PARTI	O6bg80	paramecium	206	38	62..3	889	2	Q5QU55_IDILO	Q5qu55	idiomarina
134	38	62..3	468	2	Q3EUK9_BACTE	O3euk9	bacillus th	207	38	62..3	889	2	Q3TQ15_MOUSE	Q3tql5	mus musculu
135	38	62..3	487	2	Q9PE64_NEUCR	Q9pe64	neurospora	208	38	62..3	890	1	ACON1_ECOLI	P25516	escherichia
136	38	62..3	506	2	Q2PYZ9_GLOWR	Q2pyz9	glossina mo	209	38	62..3	890	2	Q2S782_9GAMM	Q2s782	hahella che
137	38	62..3	512	2	Q3M7D8_ANAVT	Q3m7d8	anabaena va	210	38	62..3	890	2	Q37ZX7_SPHAR	Q37zx7	novosphingo
138	38	62..3	514	2	Q8GN94_SERMA	O8gn94	serratia ma	211	38	62..3	890	2	Q66A17_YERPS	O66a17	yersinia ps
139	38	62..3	535	1	HPSE_MOUSE	O6ygz1	mus musculu	212	38	62..3	890	2	Q67JK4_SYMTH	Q67jk4	syntbiobacte
140	38	62..3	550	2	Q59Y58_CANAL	O59yb8	candida alb	213	38	62..3	890	2	Q6D507_ERWCT	Q6d507	erwinia car
141	38	62..3	550	2	Q59YB4_CANAL	O59yb4	candida alb	214	38	62..3	890	2	Q7N201_CHRW	Q7nz01	chromobacte
142	38	62..3	559	2	Q4YH08_PLABE	O4yhq8	plasmodium	215	38	62..3	890	2	Q83B05_COXBU	Q83b05	coxiella bu
143	38	62..3	616	1	ACOC_SOLTU	O04916	solanum tub	216	38	62..3	890	2	Q82EF1_YERPE	O82ef1	yersinia pe
144	38	62..3	636	2	Q9Y1L4_9CILI	Q9y1l4	eufolliculi	217	38	62..3	890	2	Q3YMK9_BRARE	Q3ymk9	brachydanio
145	38	62..3	656	2	Q9BLI3_BOMMO	Q9bli3	bombyx mori	218	38	62..3	891	1	ACON_LEGPH	P37032	legionella
146	38	62..3	703	2	Q4YWP2_PLABE	Q4ywp2	plasmodium	219	38	62..3	891	2	Q95UT2_MANSE	Q95ut2	manduca sex
147	38	62..3	732	1	GP149_MOUSE	Q3uvy1	mus musculu	220	38	62..3	891	2	Q2RNJ0_RHORU	Q2rnj0	rhodospiril
148	38	62..3	741	2	Q4EDB6_9RICK	Q4edb6	wolbachia e	221	38	62..3	891	2	Q3VCW4_9SPHN	Q3vcw4	sphingopyxi
149	38	62..3	748	2	Q9B3D8_NEUCR	Q9b3d8	neurospora	222	38	62..3	891	2	Q3Z126_SHISS	Q3z126	shigella so
150	38	62..3	753	2	Q41Y86_AZOVI	Q9p3d8	neurospora	223	38	62..3	891	2	Q2N819_9SPHN	Q2n819	erythrobact
151	38	62..3	781	2	Q59F10_HUMAN	Q41y86	azotobacter	224	38	62..3	891	2	Q57N08_SALCH	Q57nu8	salmonella
152	38	62..3	790	2	Q9HBB2_HUMAN	Q59f10	homo sapien	225	38	62..3	891	2	Q5PCY9_SALPA	Q5pcy9	salmonella
153	38	62..3	806	2	Q6P289_MOUSE	Q9hbb2	homo sapien	226	38	62..3	891	2	Q5WM07_LEGPL	Q5wm07	legionella
154	38	62..3	822	2	Q9N1Q9_TRYBB	O6p289	mus musculu	227	38	62..3	891	2	Q5X417_LEGPA	Q5x417	legionella
155	38	62..3	831	2	Q8U489_PYRFU	Q9n1q9	trypanosoma	228	38	62..3	891	2	Q7N4B7_PHOIL	Q7n4b7	photorhabdu
156	38	62..3	833	2	Q6LOI2_PICTO	O8u489	pyrococcus	229	38	62..3	891	2	Q8X7C7_ECO57	Q8x7c7	escherichia
157	38	62..3	848	2	Q4J9H0_SULAC	O6l0i2	picrophilus	230	38	62..3	891	2	Q8ZP52_SALTY	Q8zps2	salmonella
158	38	62..3	848	2	Q9RAT4_LACIA	Q4j9h0	sulfolobus	231	38	62..3	891	2	Q8ZHU6_ECOL6	Q8zhu6	escherichia
159	38	62..3	848	2	Q9CHQ5_LACIA	Q9rat4	lactococcus	232	38	62..3	891	2	Q821B2_SHIFL	Q821b2	shigella fl
160	38	62..3	852	2	Q4E877_9RICK	O9chq5	lactococcus	233	38	62..3	891	2	Q827D2_SALTI	Q827d2	salmonella
161	38	62..3	853	2	Q4DEL8_TRYCR	O4e877	wolbachia e	234	38	62..3	891	2	Q7ZX37_XENLA	Q7zx37	xenopus lae
162	38	62..3	854	2	Q5YB86_RAT	Q4del8	trypanosoma	235	38	62..3	894	2	Q54X73_DICDI	Q54x73	dictyosteli
163	38	62..3	854	2	Q5YB86_RAT	Q3pi21	paracoccus	236	38	62..3	894	2	Q51W57_STRVR	Q51w57	streptomyce
164	38	62..3	855	2	Q973R5_SULTO	O5yb86	rattus norv	237	38	62..3	894	2	Q6MIH9_EDEBA	Q6mih9	bdellovibri
165	38	62..3	855	2	Q97242_SULTO	Q973r5	sulfolobus	238	38	62..3	894	2	Q7WB80_BORPA	Q7wb80	bordecella
166	38	62..3	861	2	Q731Q1_WOLPM	O97242	sulfolobus	239	38	62..3	894	2	Q7WMP8_BORBR	Q7wmp8	bordecella
167	38	62..3	862	2	Q5TEK9_HUMAN	Q731q1	wolbachia p	240	38	62..3	894	2	Q4T087_TETNG	Q4t087	tetraodon n
168	38	62..3	864	1	ITCH_MOUSE	O5tek9	homo sapien	241	38	62..3	895	2	Q41W32_AZOVI	Q41w32	azotobacter
169	38	62..3	867	2	Q5GS96_WOLTR	Q8c63	mus musculu	242	38	62..3	895	2	Q7FS8_BRUAB	Q7fs8	brucella ab
170	38	62..3	869	2	Q9HLW3_THERAC	O5gs96	wolbachia s	243	38	62..3	895	2	Q6G0V8_BRARQ	Q6gv08	bartonella
171	38	62..3	870	2	Q9YB10_AERPE	Q9hlw3	thermoplasm	244	38	62..3	895	2	Q6G528_BARHE	Q6g528	bartonella
172	38	62..3	871	2	Q5LOA4_GEOKA	O9yb10	aeropyrum p	245	38	62..3	895	2	Q8G359_BRUSU	Q8g359	brucella su
173	38	62..3	874	2	Q8ZX30_PYRAE	O5loa4	geobacillus	246	38	62..3	895	2	Q8YEM3_BRUME	Q8yem3	brucella me
174	38	62..3	875	2	Q5HA89_EHRRW	O8zx30	pyrobaculum	247	38	62..3	895	2	Q2YNU1_BRUAR	Q2ynul	brucella ab
175	38	62..3	876	2	Q571M5_MOUSE	O5ha89	ehrlichia r	248	38	62..3	895	2	Q8JVB6_9VIR2	Q8jvb6	helminthosp
176	38	62..3	877	2	Q43XD2_SOLUS	O571m5	mus musculu	249	38	62..3	896	2	Q4QDZ1_LEISHA	Q4qdz1	leishmania
177	38	62..3	878	1	ACON_BTCCN	Q43xd2	solibacter	250	38	62..3	896	2	Q2K397_RHIET	Q2k397	rhizobium e

251	38	62.3	896	2	Q5LR20_SILPO	Q5LR20_silicibacte	324	38	62.3	905	2	Q5WG13_BACSK	Q5wg13 bacillus cl
252	38	62.3	896	2	Q92L56_RHIME	Q92l56 rhizobium m	325	38	62.3	905	2	Q62AG4_BURMA	Q62ag4 burkholderi
253	38	62.3	896	2	Q98EA3_RHILO	Q98ea3 rhizobium l	326	38	62.3	905	2	Q63JU0_BURPS	Q63ju0 burkholderi
254	38	62.3	897	2	Q388J9_9TRYP	Q388j9 trypanosoma	327	38	62.3	905	2	Q6NDV4_RHOPA	Q6nda7 rhodopseude
255	38	62.3	897	2	Q5FRA9_GLUOX	Q5fra9 gluconobact	328	38	62.3	906	1	Q82KV4_STRAW	Q82kv4 streptomyce
256	38	62.3	897	2	Q8UC13_AGR75	Q8uc13 agrobacteri	329	38	62.3	906	1	ACON_BRAJA	Q70920 bradyrhizob
257	38	62.3	898	1	ACOC_ARATH	Q42560 arabidopsis	330	38	62.3	906	2	Q347Q7_RHOPA	Q347q7 rhodopseude
258	38	62.3	898	1	ACOC_CUCWA	P49608 cucurbita m	331	38	62.3	906	2	Q3VTN0_PROAE	Q3vtn0 prothecoch
259	38	62.3	898	2	Q4E5G5_TYCR	Q4e5g5 trypanosoma	332	38	62.3	906	2	Q4H891_9DEIO	Q4h891 deinococcus
260	38	62.3	898	2	Q81396_CITLI	Q81396 citrus limo	333	38	62.3	906	2	Q9RTN7_DEIRA	Q9rtn7 deinococcus
261	38	62.3	898	2	Q6YZX6_ORISA	Q6yzx6 oryza sativ	334	38	62.3	907	2	Q9SZ36_ARATH	Q9sz36 arabidopsis
262	38	62.3	898	2	Q84NI5_LYCPN	Q84ni5 lycopersico	335	38	62.3	907	2	Q3CWY9_ALTAT	Q3cwy9 pseudalter
263	38	62.3	898	2	Q84TR4_LYCPN	Q84tr4 lycopersico	336	38	62.3	907	2	Q3JA87_NITOC	Q3ja87 nitrosococc
264	38	62.3	898	2	Q9FV89_TOBAC	Q9fve9 nicotiana t	337	38	62.3	907	2	Q41DJ0_9BACI	Q41dj0 exiguobacte
265	38	62.3	898	2	Q8SIB9_ARATH	Q8sib9 arabidopsis	338	38	62.3	907	2	Q4MNH3_BACCE	Q4mnh3 bacillus ce
266	38	62.3	899	2	Q76935_DROME	Q76935 drosophila	339	38	62.3	907	2	Q637L2_BACCK	Q637l2 bacillus th
267	38	62.3	899	2	Q9NFX2_DROME	Q9nfx2 drosophila	340	38	62.3	907	2	Q6HFI4_BACHK	Q6hfi4 bacillus th
268	38	62.3	899	2	Q9VG23_DROME	Q9vg23 drosophila	341	38	62.3	907	2	Q81AF6_BACCR	Q81af6 bacillus th
269	38	62.3	899	2	Q8G4I5_BIFLO	Q8g4i5 bifidobacte	342	38	62.3	907	2	Q81V80_BACAN	Q81v80 bacillus an
270	38	62.3	899	2	Q8VDC3_MOUSE	Q8vdc3 mus musculu	343	38	62.3	907	2	Q9KAI8_BACHD	Q9ka18 bacillus ha
271	38	62.3	900	2	Q4EG12_LISMO	Q4eg12 listeria mo	344	38	62.3	907	2	Q733M3_BACCI	Q733m3 bacillus ce
272	38	62.3	900	2	Q4EPF1_LISMO	Q4epf1 listeria mo	345	38	62.3	908	1	ACON_BACSU	P09339 bacillus su
273	38	62.3	900	2	Q71Z28_LISMF	Q71z28 listeria mo	346	38	62.3	908	2	Q3NTJ4_SHEFR	Q3ntj4 shewanella
274	38	62.3	900	2	Q8Y6P3_LISMO	Q8y6p3 listeria mo	347	38	62.3	908	2	Q3RAW9_XYLFA	Q3raw9 xyella fas
275	38	62.3	900	2	Q92B68_LISIN	Q92b68 listeria in	348	38	62.3	908	2	Q3RB71_XYLFA	Q3rb71 xyella fas
276	38	62.3	901	1	ACON_STAAC	Q5hg63 staphylococ	349	38	62.3	908	2	Q3RER2_XYLFA	Q3rer2 xyella fas
277	38	62.3	901	1	ACON_STRAA	P63433 staphylococ	350	38	62.3	908	2	Q62UI4_BACLD	Q62ui4 bacillus li
278	38	62.3	901	1	ACON_STRAA	P99148 staphylococ	351	38	62.3	908	2	Q65J27_BACLD	Q65j27 bacillus li
279	38	62.3	901	1	ACON_STRAA	Q6gh55 staphylococ	352	38	62.3	908	2	Q9PGK8_XYLFA	Q9pgk8 xyella fas
280	38	62.3	901	1	ACON_STRAA	Q6g9k9 staphylococ	353	38	62.3	910	2	Q2SIF3_98PHI	Q2sif3 salinibacte
281	38	62.3	901	1	ACON_STRAA	P63434 staphylococ	354	38	62.3	910	2	Q44G01_CHRSL	Q44g01 chromohalob
282	38	62.3	901	1	ACON_STRAQ	Q5hpj0 staphylococ	355	38	62.3	910	2	Q913F5_PSEAE	Q913f5 pseudomonas
283	38	62.3	901	1	ACON_STRAE	Q6cp22 staphylococ	356	38	62.3	911	2	Q4NRQ5_CHLLI	Q4nrq5 chlorobium
284	38	62.3	901	2	Q6SYX7_AEDAE	Q6syc7 aedes aegyp	357	38	62.3	912	2	Q4N6X0_THEPA	Q4n6x0 theileria p
285	38	62.3	901	2	Q7PTD5_ANOAG	Q7ptd5 anopheles g	358	38	62.3	913	2	Q2XE12_PSEPU	Q2xe12 pseudomonas
286	38	62.3	901	2	Q3S1E3_RALME	Q3s1e3 ralstonia m	359	38	62.3	913	2	Q3DYH8_CHLPU	Q3dyh8 chloroflexu
287	38	62.3	901	2	Q46YU0_RALEJ	Q46yu0 ralstonia e	360	38	62.3	913	2	Q3K9C7_PSEPP	Q3k9c7 pseudomonas
288	38	62.3	901	2	Q2YXW7_STAAB	Q2yxw7 staphylococ	361	38	62.3	913	2	Q4KFD5_PSEFS	Q4kfd5 pseudomonas
289	38	62.3	901	2	Q2KXA5_BORAV	Q2kxa5 bordetella	362	38	62.3	913	2	Q8BL24_PSEPK	Q8bl24 pseudomonas
290	38	62.3	901	2	Q4L658_STAHT	Q4l658 staphylococ	363	38	62.3	914	2	Q353T3_9GAMM	Q353t3 alkaliilmi
291	38	62.3	901	2	Q5P9X0_ANAMJ	Q5p9x0 anaplasma m	364	38	62.3	914	2	Q31FL7_PSEHT	Q31fl7 pseudoalter
292	38	62.3	901	2	Q7UMD2_RHOBA	Q7umd2 rhodopirell	365	38	62.3	914	2	Q3J5I1_RHOSA	Q3j5i1 rhodobacter
293	38	62.3	901	2	Q7VX12_BORPE	Q7vx12 bordetella	366	38	62.3	914	2	Q48GJ9_PSE14	Q48gj9 pseudomonas
294	38	62.3	901	2	Q7W7M5_BORPA	Q7w7m5 bordetella	367	38	62.3	914	2	Q4ZQY6_PSEU2	Q4zqy6 pseudomonas
295	38	62.3	901	2	Q7WLA5_BORBR	Q7wla5 bordetella	368	38	62.3	914	2	Q884S2_PSESM	Q884s2 pseudomonas
296	38	62.3	901	2	Q8XXW0_RALSO	Q8xxw0 ralstonia s	369	38	62.3	915	2	Q3QVD5_9RHOB	Q3qvd5 silicibacte
297	38	62.3	902	2	Q76934_DROME	Q76934 drosophila	370	38	62.3	915	2	Q4APY6_9CHLB	Q4apy6 chlorobium
298	38	62.3	902	2	Q9NFX3_DROME	Q9nfx3 drosophila	371	38	62.3	916	2	Q47NLI_THEFY	Q47nli thermobifid
299	38	62.3	902	2	Q9VCV4_DROME	Q9vcv4 drosophila	372	38	62.3	917	2	Q3LS03_NATPD	Q3ls03 natronomona
300	38	62.3	902	2	Q6YNR9_PRUAV	Q6ynr9 prunus aviu	373	38	62.3	917	2	Q6F827_ACIAD	Q6f827 acinetobact
301	38	62.3	902	2	Q49XB4_STAS1	Q49xe4 staphylococ	374	38	62.3	919	2	Q3P888_PARDE	Q3p888 paracoccus
302	38	62.3	902	2	Q5SMF6_THET8	Q5smf6 thermus the	375	38	62.3	920	2	Q36TD9_MARHY	Q36td9 marinobacte
303	38	62.3	902	2	Q72KN3_THET2	Q72kn3 thermus the	376	38	62.3	920	2	Q2J3F8_RHOPA	Q2j3f8 rhodopseude
304	38	62.3	903	1	ITCH_HUMAN	Q96j02 homo sapien	377	38	62.3	922	2	Q50579_XANCA	Q50579 xanthomonas
305	38	62.3	903	2	Q2WL20_MAGSA	Q2wl20 magnetospir	378	38	62.3	922	2	Q3BUA8_XANCS	Q3bua8 xanthomonas
306	38	62.3	903	2	Q8EQL5_OCEIH	Q8eq15 oceanobacil	379	38	62.3	922	2	Q43G24_9CHLB	Q43g24 chlorobium
307	38	62.3	903	2	Q9A299_CAUCR	Q9a299 caulobacter	380	38	62.3	922	2	Q4UU91_XANCS	Q4uu91 xanthomonas
308	38	62.3	904	2	Q76BK5_ACEAC	Q76bk5 acetobacter	381	38	62.3	922	2	Q2P1V2_XANOR	Q2p1v2 xanthomonas
309	38	62.3	904	2	Q9RNH9_STRCO	Q9rn9 streptomyce	382	38	62.3	922	2	Q5GYV2_XANOR	Q5gyv2 xanthomonas
310	38	62.3	905	2	Q2T7J6_BURTH	Q2t7j6 burkholderi	383	38	62.3	922	2	Q8P9K3_XANCP	Q8p9k3 xanthomonas
311	38	62.3	905	2	Q35L82_9BRAD	Q35l82 bradyrhizob	384	38	62.3	922	2	Q8PLCO_XANAC	Q8plco xanthomonas
312	38	62.3	905	2	Q37B14_RHOPA	Q37b14 rhodopseude	385	38	62.3	923	2	Q3WUN1_9RHIZ	Q3wun1 mesorhizobi
313	38	62.3	905	2	Q37K18_RHOBA	Q37k18 rhodopseude	386	38	62.3	925	2	Q87ER4_XYLFT	Q87er4 xyella fas
314	38	62.3	905	2	Q3FH02_9BURK	Q3fh02 burkholderi	387	38	62.3	927	2	Q5V7D3_HALMA	Q5v7d3 haloarcula
315	38	62.3	905	2	Q3PV01_NITWA	Q3pv01 nitrobacter	388	38	62.3	929	2	Q4UIP2_THEAN	Q4ui2 theileria a
316	38	62.3	905	2	Q3SVV1_NITWN	Q3sevl nitrobacter	389	38	62.3	931	2	Q405H4_9RHOB	Q405h4 jannaschia
317	38	62.3	905	2	Q44XY7_9BURK	Q44xy7 burkholderi	390	38	62.3	931	2	Q9RILI_STVRV	Q9rili streptomyc
318	38	62.3	905	2	Q4BA86_BURVI	Q4ba86 burkholderi	391	38	62.3	933	2	Q5YU10_NOCFA	Q5yu10 nocardia fa
319	38	62.3	905	2	Q4LR23_9BURK	Q4lr23 burkholderi	392	38	62.3	934	1	ACON_CORDI	Q6nh63 corynebacte
320	38	62.3	905	2	Q393U7_BUR33	Q393u7 burkholderi	393	38	62.3	934	2	Q46EK0_METBA	Q46ek0 methanocarc
321	38	62.3	905	2	Q3JKE4_BURP1	Q3jke4 burkholderi	394	38	62.3	935	2	Q3GS30_9GAMM	Q3gs30 psychrobact
322	38	62.3	905	2	Q2KXP1_BORAV	Q2kxp1 bordetella	395	38	62.3	935	2	Q74EW3_GEOSL	Q74ew3 geobacter s
323	38	62.3	905	2	Q5P0Q2_AZOSE	Q5p0q2 azoarcus sp	396	38	62.3	936	1	ACON_CORJK	Q4jvm4 corynebacte

397	38	62.3	936	2	Q4NAY6_9M1CC	Q4nay6 arthrobacte	470	37	60.7	1470	2	Q89255_9POTY	Q89255 wheat spind
398	38	62.3	937	1	ACON_COREF	Q8fta8 corynebacte	471	37	60.7	1608	2	Q4SHG8_TETNG	Q4shg8 tetraodon n
399	38	62.3	937	2	Q39WM6_GEOMG	Q39wm6 geobacter m	472	37	60.7	1968	2	Q2Q403_MOUSE	Q2q403 mus musculus
400	38	62.3	937	2	Q5N111_FRAVT	Q5n111 francisella	473	37	60.7	1969	2	Q2PZW9_MOUSE	Q2pzw9 mus musculus
401	38	62.3	937	2	Q5N165_AZOSE	Q5ny65 azoarcus sp	474	37	60.7	2014	2	Q2Q404_MOUSE	Q2q404 mus musculus
402	38	62.3	938	2	Q3GTD5_9ACTO	Q3gtd5 nocardioide	475	37	60.7	2160	2	Q3U1K8_DROME	Q3u1k8 drosophila
403	38	62.3	939	1	ACON_CORGL	Q8nq98 corynebacte	476	37	60.7	2178	2	Q3W4X4_DROME	Q3w4x4 drosophila
404	38	62.3	939	2	Q95V48_DROME	Q95v48 drosophila	477	37	60.7	2193	2	Q3R3M6_HUMAN	Q3r3m6 homo sapien
405	38	62.3	939	2	Q4FQ00_PSYAR	Q4fq00 psychrobact	478	37	60.7	2266	2	Q2PZW8_HUMAN	Q2pzw8 homo sapien
406	38	62.3	939	2	Q3SKE1_THIDA	Q3ske1 thiobacillu	479	37	60.7	2365	2	Q2Q3Z9_HUMAN	Q2q3z9 homo sapien
407	38	62.3	940	2	Q8TU24_METAC	Q8tu24 methanosarc	480	37	60.7	2491	2	Q4YN06_PLABE	Q4ynq6 plasmodium
408	38	62.3	942	2	Q8PWP8_METMA	Q8pwp8 methanosarc	481	37	60.7	2598	2	Q7RNQ6_PLAYO	Q7rnq6 plasmodium
409	38	62.3	943	2	Q53166_MYCTU	Q53166 mycobacteri	482	37	60.7	2860	2	Q8I640_PLAF7	Q8i640 plasmodium
410	38	62.3	943	2	Q7U000_MYCBO	Q7u000 mycobacteri	483	36.5	59.8	143	2	Q9L9G4_STRSH	Q9l9g4 streptomyc
411	38	62.3	944	2	Q8CBL3_MYCLE	Q8cbl3 mycobacteri	484	36.5	59.8	455	2	Q4K132_PSEF5	Q4k132 pseudomonas
412	38	62.3	945	2	Q2IFP5_9DELT	Q2ifp5 anaeromyxob	485	36.5	59.8	465	2	Q3KHV0_PSEPF	Q3khv0 pseudomonas
413	38	62.3	946	2	Q2YAP3_NITWU	Q2yap3 nitrospirir	486	36	59.0	82	2	Q43TF6_SOLUS	Q43tf6 solidibacter
414	38	62.3	947	2	Q3ND29_9PROT	Q3nd29 nitrosomona	487	36	59.0	89	1	RS15_METCA	Q609c5 methylococ
415	38	62.3	948	2	Q6MCW0_PARUM	Q6mcw0 parachlamyd	488	36	59.0	111	2	Q3BFR0_PLAFA	Q3bfr0 plasmodium
416	38	62.3	948	2	Q82VR3_NITRU	Q82vr3 nitrosomona	489	36	59.0	111	2	Q3BI67_PLAFA	Q3bi67 plasmodium
417	38	62.3	949	2	Q40ZB2_KINRA	Q40zb2 kineococcus	490	36	59.0	116	2	Q3BFV2_PLAFA	Q3bfv2 plasmodium
418	38	62.3	959	2	Q740Z1_MYCPA	Q740z1 mycobacteri	491	36	59.0	119	2	Q3BI84_PLAFA	Q3bie4 plasmodium
419	38	62.3	960	2	Q6AFD6_LEIIX	Q6afd6 leifsonia x	492	36	59.0	120	2	Q3BIY3_PLAFA	Q3biy3 plasmodium
420	38	62.3	961	1	ACON_MYCAV	Q8451 mycobacteri	493	36	59.0	120	2	Q3BIY4_PLAFA	Q3biy4 plasmodium
421	38	62.3	966	2	Q4BS12_9BURK	Q4bs12 polaromonas	494	36	59.0	120	2	Q3BJF9_PLAFA	Q3bjf9 plasmodium
422	38	62.3	978	2	Q3FMD1_9BURK	Q3fmd1 rhodiferax	495	36	59.0	122	2	Q3BG74_PLAFA	Q3bg74 plasmodium
423	38	62.3	984	2	Q3PNA3_NITRA	Q3pna3 nitrobacter	496	36	59.0	123	2	Q89YA2_BRAJA	Q89ya2 bradyrhizob
424	38	62.3	980	2	Q8L784_ARATH	Q8l784 arabidopsis	497	36	59.0	129	2	Q3BEX6_PLAFA	Q3bex6 plasmodium
425	38	62.3	995	2	Q94A28_ARATH	Q94a28 arabidopsis	498	36	59.0	129	2	Q3BHQ0_PLAFA	Q3bhq0 plasmodium
426	38	62.3	1051	2	Q483R8_COLP3	Q483r8 colwellia p	499	36	59.0	167	2	Q2SP07_9GAMM	Q2sp07 habella che
427	38	62.3	1052	2	Q7MZE8_PHOLL	Q7mze8 photorhabdu	500	36	59.0	175	2	Q4L5S2_STAHI	Q4l5s2 staphylococ
428	38	62.3	1052	2	Q6DUK1_TOXGO	Q6duk1 toxoplasma	501	36	59.0	180	2	Q5QDR6_9HIV1	Q5qdr6 human immun
429	38	62.3	1278	2	Q3X316_9ACTN	Q3x316 rubrobacter	502	36	59.0	185	2	Q5XBU0_STRP6	Q5xbu0 streptococ
430	38	62.3	2568	2	Q2UN49_9SPOR	Q2un49 aspergillus	503	36	59.0	185	2	Q8P0R5_STRP8	Q8p0r5 streptococ
431	37.5	61.5	1427	2	Q339H4_ORYSA	Q339h4 oryza sativ	504	36	59.0	188	2	Q74KQ4_LACJO	Q74kq4 lactobacill
432	37	60.7	68	2	Q5L3X1_GEOKA	Q5l3x1 geobacillus	505	36	59.0	189	2	Q4IHW7_GIBZE	Q4ihw7 gibberella
433	37	60.7	109	2	Q9NHJ2_PLAFA	Q9nhj2 plasmodium	506	36	59.0	189	2	Q8WT06_PLAFA	Q8wtd6 plasmodium
434	37	60.7	117	2	Q8TX48_PLAFA	Q8tx48 plasmodium	507	36	59.0	191	2	Q6CE95_YARLI	Q6ce95 yarrowia li
435	37	60.7	118	2	Q9X297_PLAFA	Q9x297 plasmodium	508	36	59.0	206	2	Q6MJ72_BDEBA	Q6mj72 bdellovibri
436	37	60.7	120	2	Q3BHD7_PLAFA	Q3bhd7 plasmodium	509	36	59.0	210	2	Q9MC50_BPD3	Q9mc50 bacterioph
437	37	60.7	120	2	Q3BHD8_PLAFA	Q3bhd8 plasmodium	510	36	59.0	212	2	Q58M91_9CAUD	Q58m91 cyanophaga
438	37	60.7	122	2	Q3BFY8_PLAFA	Q3bfy8 plasmodium	511	36	59.0	218	2	Q83WZ5_STRRO	Q83wz5 streptomyc
439	37	60.7	157	2	Q6NKK5_CORDI	Q6nkk5 corynebacte	512	36	59.0	236	2	Q4KME2_BRARE	Q4kme2 brachydanio
440	37	60.7	168	2	Q9U7K3_PLAFA	Q9u7k3 plasmodium	513	36	59.0	238	2	Q6DEM7_BRARE	Q6dem7 brachydanio
441	37	60.7	170	2	Q4YEH6_PLABE	Q4yeh6 plasmodium	514	36	59.0	269	2	Q3CUP1_ALTAT	Q3cup1 pseudoalter
442	37	60.7	176	2	Q3X353_9ACTN	Q3x353 rubrobacter	515	36	59.0	306	2	Q4HJ79_CAMLA	Q4hj79 campylobact
443	37	60.7	177	2	Q5PJ40_SALPA	Q5pj40 salmonella	516	36	59.0	318	2	Q6DI37_BRARE	Q6di37 brachydanio
444	37	60.7	201	2	Q8YS91_ANASP	Q8ys91 anabaena sp	517	36	59.0	336	2	Q7Q4X5_ANOGA	Q7q4x5 anopheles g
445	37	60.7	202	2	Q4C6U7_CROWT	Q4c6u7 crocosphaer	518	36	59.0	341	2	Q8LEJ4_ARATH	Q8lej4 arabidopsis
446	37	60.7	238	2	Q40PP5_DESAC	Q40pp5 desulfuromo	519	36	59.0	341	2	Q9FLV0_ARATH	Q9flv0 arabidopsis
447	37	60.7	237	2	Q8V984_9GEMI	Q8v984 soybean cri	520	36	59.0	345	2	Q3YSM8_EHRCJ	Q3ysm8 ehrlichia c
448	37	60.7	323	2	Q9XIS2_ARATH	Q9xib2 arabidopsis	521	36	59.0	378	2	Q4RRW7_TETNG	Q4rrw7 tetraodon n
449	37	60.7	328	2	Q54PM4_DICDI	Q54pm4 dictyostell	522	36	59.0	393	2	Q2U774_ASPOR	Q2u774 aspergillus
450	37	60.7	354	2	Q5ULH1_9CAUDI	Q5ulh1 lactobacill	523	36	59.0	400	2	Q48SH4_STRPM	Q48sh4 streptococ
451	37	60.7	383	2	Q3I629_9ANNE	Q3i629 aporetcode	524	36	59.0	400	2	Q938J4_9CAUD	Q938j4 temperate p
452	37	60.7	383	2	Q3I701_9ANNE	Q3i701 aporetcode	525	36	59.0	400	2	Q79XS7_STRP3	Q79xs7 streptococ
453	37	60.7	384	2	Q3I627_9ANNE	Q3i627 dendrobaena	526	36	59.0	400	2	Q5XEG3_STRP6	Q5xeg3 streptococ
454	37	60.7	384	2	Q77072_EISFO	Q77072 eisenia foe	527	36	59.0	401	2	Q5XAT5_STRP6	Q5xat5 streptococ
455	37	60.7	385	2	Q3I628_9ANNE	Q3i628 aporetcode	528	36	59.0	401	2	Q8NZT3_STRP8	Q8nzt3 streptococ
456	37	60.7	449	2	Q98KY4_RHILO	Q98ky4 rhizobium l	529	36	59.0	402	2	Q53U85_ASPOR	Q53u85 aspergillus
457	37	60.7	661	2	Q6L889_MOUSE	Q6l889 mus musculu	530	36	59.0	402	2	Q8K797_STRP3	Q8k797 streptococ
458	37	60.7	662	2	Q9WV57_RAT	Q9wv57 rattus norv	531	36	59.0	402	2	Q8K7N4_STRP3	Q8k7n4 streptococ
459	37	60.7	721	2	Q52KF9_MOUSE	Q52kf9 mus musculu	532	36	59.0	409	2	Q8GMW0_9STRE	Q8gmw0 streptococ
460	37	60.7	722	2	Q3X5P8_9ACTN	Q3x5p8 rubrobacter	533	36	59.0	425	2	Q96517_ENTCL	Q96517 enterobacte
461	37	60.7	730	1	GP149_RAT	Q924y8 rattus norv	534	36	59.0	433	2	Q2ZGQ2_CALUS	Q2zgg2 caldicellul
462	37	60.7	756	1	NAO2_HAETN	P43837 haemophilus	535	36	59.0	437	2	Q4CEP6_CLOTM	Q4cep6 clostridum
463	37	60.7	757	2	Q4QJX2_HAEI8	Q4qjx2 haemophilus	536	36	59.0	442	2	Q6E5B1_9AGAR	Q6esb1 volvoxella
464	37	60.7	807	2	Q5CG31_CRYHO	Q5cg31 cryptospori	537	36	59.0	451	1	V110_FOWPV	P36700 fowlpox vir
465	37	60.7	980	2	Q2Q402_MOUSE	Q2q402 mus musculu	538	36	59.0	451	2	Q70H45_FOWPV	Q70h45 mycoplasma
466	37	60.7	1109	2	Q18628_CABEL	Q18628 caenorhabdi	539	36	59.0	470	2	Q601N5_MWCH2	Q601n5 mycoplasma
467	37	60.7	1144	2	Q2Q400_HUMAN	Q2q400 homo sapien	540	36	59.0	523	1	HPSE_CHICK	Q90yk5 gallus gall
468	37	60.7	1260	1	YAOE_SCHPO	Q10093 schizosacch	541	36	59.0	524	2	Q7T198_CHICK	Q7t198 gallus gall
469	37	60.7	1308	2	Q8MYR7_DROME	Q8myr7 drosophila	542	36	59.0	545	2	Q5NMW3_ZYMMO	Q5nmw3 zymomonas m



543	36	59.0	553	2	Q5C1S9_CRYHO	Q5C1S9_cryptospori	616	35	57.4	437	2	Q35BZ0_9BRAD	Q35BZ0_bradyrhizob
544	36	59.0	553	2	Q5C0S3_CRYPV	Q5C0S3_cryptospori	617	35	57.4	456	2	Q5CW59_XANOR	Q5CW59_xanthomonas
545	36	59.0	577	2	Q3K3M3_9NOST	Q3K3M3_anabaena ci	618	35	57.4	458	2	Q6MRP9_BDEBA	Q6MRP9_bdellovibri
546	36	59.0	585	2	Q3US71_MOUSE	Q3US71_mus musculus	619	35	57.4	460	2	P72256_RHOER	P72256_rhodococcus
547	36	59.0	604	2	Q98R71_MYCPU	Q98R71_mycoplasma	620	35	57.4	460	2	P52819_RHOOP	P52819_rhodococcus
548	36	59.0	622	2	Q2ZJU5_CALSA	Q2ZJU5_caldicellul	621	35	57.4	460	2	Q8VQU6_9NOCA	Q8VQU6_rhodococcus
549	36	59.0	622	2	Q5NQM1_ZYMO	Q5NQM1_zymomonas m	622	35	57.4	497	2	Q7STO6_9HIV1	Q7STO6_human immun
550	36	59.0	634	2	Q5ONU9_ENTHI	Q5ONU9_entamoeba h	623	35	57.4	497	2	Q7STO7_9HIV1	Q7STO7_human immun
551	36	59.0	634	2	Q5OS46_ENTHI	Q5OS46_entamoeba h	624	35	57.4	497	2	Q7STO8_9HIV1	Q7STO8_human immun
552	36	59.0	637	2	Q4SRP7_DICDI	Q4SRP7_dictyosteli	625	35	57.4	497	2	Q7STO9_9HIV1	Q7STO9_human immun
553	36	59.0	637	2	Q5ENW7_9VIRU	Q5ENW7_torque teno	626	35	57.4	501	2	Q3V7H9_DESUF	Q3V7H9_desulfotale
554	36	59.0	656	2	Q59G95_HUMAN	Q59G95_homo sapien	627	35	57.4	511	2	Q5KCC3_CRYNE	Q5KCC3_cryptococcu
555	36	59.0	657	2	Q840A5_9STRE	Q840A5_streptococc	628	35	57.4	511	2	Q46L37_PROMT	Q46L37_prochloroco
556	36	59.0	680	2	Q8TS98_METAC	Q8TS98_methanosarc	629	35	57.4	515	2	Q87KB5_VIBPA	Q87KB5_vibrio para
557	36	59.0	700	2	Q8GCM8_MYCFE	Q8GCM8_mycoplasma	630	35	57.4	518	2	Q4ZSF3_PSEU2	Q4ZSF3_pseudomonas
558	36	59.0	728	1	ARPA_ECOLI	P23325 escherichia	631	35	57.4	519	2	Q5SJD6_CRYNE	Q5SJD6_cryptococcu
559	36	59.0	728	2	Q2M6T7_ECOLI	Q2M6T7_escherichia	632	35	57.4	519	2	Q2XKB4_PSRPK	Q2XKB4_pseudomonas
560	36	59.0	728	2	Q8X606_ECO57	Q8X606_escherichia	633	35	57.4	530	2	Q88HB4_PSRPK	Q88HB4_pseudomonas
561	36	59.0	833	2	Q3BP07_HALRO	Q3BP07_halocynthia	634	35	57.4	543	2	Q5KDZ5_CRYNE	Q5KDZ5_cryptococcu
562	36	59.0	865	1	AP1G1_SCHPO	Q9UPG7 schizosacch	635	35	57.4	565	2	Q381E9_9TRYP	Q381E9_trypanosoma
563	36	59.0	915	1	MGR7_HUMAN	Q14831_homo sapien	636	35	57.4	581	2	Q4S6T6_TETNG	Q4S6T6_tetradon n
564	36	59.0	915	2	Q9BDT8_TENEC	Q9BDT8_tenrec ecau	637	35	57.4	602	2	Q5TTI3_ANOGA	Q5TTI3_anopheles g
565	36	59.0	922	1	MGR7_PONPY	Q5Rdq8 pongo pygma	638	35	57.4	610	1	C4BP_BOVIN	Q28065_bos taurus
566	36	59.0	930	2	Q3BDT5_9RUTH	Q3BDT5_rhynchocyon	639	35	57.4	610	1	Q2KJ87_BOVIN	Q2KJ87_bos taurus
567	36	59.0	985	2	Q9XSW7_PIG	Q9XSW7_sus scrofa	640	35	57.4	625	1	ADEC2_ERAJA	Q89H53_bradyrhizob
568	36	59.0	1028	1	CNTN3_MOUSE	Q07409_mus musculus	641	35	57.4	650	2	Q9SGR9_ARATH	Q9SGR9_arabidopsis
569	36	59.0	1028	1	CNTN3_RAT	Q62682_rattus norv	642	35	57.4	650	2	Q5TTL6_ANOGA	Q5TTL6_anopheles g
570	36	59.0	1055	2	Q36U69_MARHY	Q36U69_marinobacte	643	35	57.4	664	2	Q4T9K4_TETNG	Q4T9K4_tetradon n
571	36	59.0	1082	2	Q4RG51_TETNG	Q4RG51_tetradon n	644	35	57.4	666	1	EXOB4_EMENT	Q5B6X0_emicella
572	36	59.0	1513	2	Q5B955_EMENT	Q5B955_aspergillus	645	35	57.4	739	2	Q9LN30_ARATH	Q9LN30_arabidopsis
573	36	59.0	1526	2	Q2JPO1_9CIAN	Q2JPO1_cyanobacter	646	35	57.4	757	2	Q8RXQ7_ARATH	Q8RXQ7_arabidopsis
574	36	59.0	1967	2	Q7RHF3_PLAYO	Q7RHF3_plasmodium	647	35	57.4	796	2	Q3EGA2_ACTSC	Q3EGA2_actinobacil
575	36	59.0	5277	2	Q55EP8_DICDI	Q55EP8_dictyosteli	648	35	57.4	796	2	Q65V64_HANSM	Q65V64_mannheimia
576	35.5	58.2	230	2	Q55EP8_DICDI	Q55EP8_dictyosteli	649	35	57.4	813	2	Q75229_HUMAN	Q75229_homo sapien
577	35	57.4	81	2	Q3XV4_RHLIZ	Q3XV4_mesothrizobi	650	35	57.4	830	2	Q4PGF1_USTMA	Q4PGF1_ustilaco ma
578	35	57.4	88	2	Q3RG53_FUSNN	Q3RG53_fusobacteri	651	35	57.4	849	2	Q4YRM2_PLABE	Q4YRM2_plasmodium
579	35	57.4	104	2	Q941V6_ORISA	Q941V6_oryza sativ	652	35	57.4	864	2	Q59GZ9_HUMAN	Q59GZ9_homo sapien
580	35	57.4	117	2	Q9YBL9_AERPE	Q9YBL9_aeropyrum p	653	35	57.4	868	1	TOP1_HAEIN	P43012_haemophilus
581	35	57.4	140	2	Q6EUZ2_9CUCU	Q6EUZ2_curculio gl	654	35	57.4	868	2	Q4QK75_HAB18	Q4QK75_haemophilus
582	35	57.4	146	2	Q59S70_CANAL	Q59S70_candida alb	655	35	57.4	870	2	Q3EHU0_ACTSC	Q3EHU0_actinobacil
583	35	57.4	177	2	Q4WC02_ASPFU	Q4WC02_aspergillus	656	35	57.4	875	2	Q97FK6_CLOAB	Q97FK6_clostridium
584	35	57.4	178	2	Q3IHS8_PSEHT	Q3IHS8_pseudalter	657	35	57.4	914	1	PERT_MOUSE	P35419_mus musculus
585	35	57.4	209	2	Q2KBB2_RHET	Q2KBB2_rhizobium e	658	35	57.4	914	1	PERT_RAT	P14650_rattus norv
586	35	57.4	212	2	Q5YPO5_NOCFA	Q5YPO5_nocardia fa	659	35	57.4	915	1	MGR7_MOUSE	P686d2_mus musculus
587	35	57.4	251	2	Q59H96_HUMAN	Q59H96_homo sapien	660	35	57.4	915	1	MGR7_RAT	P35400_rattus norv
588	35	57.4	253	2	Q4CE46_CLOTM	Q4CE46_clostridium	661	35	57.4	916	2	Q53QT2_HUMAN	Q53QT2_homo sapien
589	35	57.4	266	2	Q7LE08_HUMAN	Q7LE08_homo sapien	662	35	57.4	926	1	PERT_PIG	P09333_sus scrofa
590	35	57.4	270	2	Q69RN3_ORISA	Q69RN3_oryza sativ	663	35	57.4	933	1	PERT_CANFA	Q8HYD7_canis famil
591	35	57.4	285	2	Q5JJI8_PYRKO	Q5JJI8_pyrococcus	664	35	57.4	933	1	PERT_HUMAN	P07202_homo sapien
592	35	57.4	316	2	Q2KWR7_BORAV	Q2KWR7_bordetella	665	35	57.4	933	2	Q502Y3_HUMAN	Q502Y3_homo sapien
593	35	57.4	317	2	Q8LPK1_ARATH	Q8LPK1_arabidopsis	666	35	57.4	936	1	PCDA5_HUMAN	Q9Y5H7_homo sapien
594	35	57.4	324	2	Q4IM30_GIBZE	Q4IM30_gibberella	667	35	57.4	936	1	PCDA5_PANTR	Q5DR7_pan troglod
595	35	57.4	324	2	Q9FJQ3_ARATH	Q9FJQ3_arabidopsis	668	35	57.4	936	2	Q5R6N6_PONPY	Q5R6N6_pongo pygma
596	35	57.4	343	2	Q31271_RAT	Q31271_rattus norv	669	35	57.4	1010	2	Q5SPG5_CRYNE	Q5SPG5_cryptococcu
597	35	57.4	349	2	Q7USP3_RHOBA	Q7USP3_rhodopirell	670	35	57.4	1208	1	RECQ4_HUMAN	Q94761_homo sapien
598	35	57.4	351	2	Q35AM3_3BRAD	Q35AM3_bradyrhizob	671	35	57.4	1215	2	P96937_MYCTU	P96937_mycobacteri
599	35	57.4	351	2	Q6SPF5_PBACT	Q6SPF5_uncultured	672	35	57.4	1215	2	Q7ULJ6_MYCBO	Q7ULJ6_mycobacteri
600	35	57.4	351	2	Q8RTU6_9PROT	Q8RTU6_uncultured	673	35	57.4	1308	2	Q8RV60_ARATH	Q8RV60_arabidopsis
601	35	57.4	361	2	Q8BKES_MOUSE	Q8BKES_mus musculus	674	35	57.4	1398	2	Q8VKH0_MYCTU	Q8VKH0_mycobacteri
602	35	57.4	369	2	Q2UFH1_ASPOR	Q2UFH1_aspergillus	675	35	57.4	1423	2	Q6FSL0_CANGA	Q6FSL0_candida gla
603	35	57.4	381	2	Q7PFI4_ANOGA	Q7PFI4_anopheles g	676	35	57.4	1793	2	Q8EWM0_MYCPE	Q8EWM0_mycoplasma
604	35	57.4	384	2	Q17962_CAEEL	Q17962_caenorhabdi	677	35	57.4	1943	2	Q4KKQ1_SCHMA	Q4KKQ1_schistosoma
605	35	57.4	402	1	PROB_SCHPO	Q1J380_schizosacch	678	35	57.4	2154	2	Q63L14_BURPS	Q63L14_burkholderi
606	35	57.4	415	1	SACB_ERWAM	Q46654_erwinia amy	679	35	57.4	2308	2	Q7N4L0_PHOLL	Q7N4L0_photorhabdu
607	35	57.4	415	2	Q4R0I7_ERWAM	Q4R0I7_erwinia amy	680	35	57.4	2473	2	Q8IEQ0_PLAF7	Q8IEQ0_plasmodium
608	35	57.4	418	2	Q6CTT1_KULLA	Q6CTT1_kluyveromyc	681	34.5	56.6	246	2	Q6DV18_9PERC	Q6DV18_rachycentro
609	35	57.4	420	2	Q75EY9_ASHGO	Q75EY9_aeshbva goas	682	34.5	56.6	368	2	Q36MI8_MARHY	Q36MI8_marinobacte
610	35	57.4	423	1	YHJ3_YEAST	P38690_saccharomyc	683	34	55.7	31	2	Q7R875_PLAYO	Q7R875_plasmodium
611	35	57.4	423	2	Q5ACK3_CANAL	Q5ACK3_candida alb	684	34	55.7	42	2	Q8FHT1_ECOL6	Q8FHT1_escherichia
612	35	57.4	423	2	Q66R86_YEAST	Q66R86_saccharomyc	685	34	55.7	68	2	Q9JE16_9VIRU	Q9JE16_tonate viru
613	35	57.4	424	2	Q6BXH8_DEBHA	Q6BXH8_debaryomyce	686	34	55.7	72	2	Q97HV4_CLOAB	Q97HV4_clostridium
614	35	57.4	428	1	PROB_YEAST	P32264_saccharomyc	687	34	55.7	116	2	Q40K87_EHRCH	Q40K87_ehrlichia c
615	35	57.4	430	1	SNX4_CANGA	Q5FPT9_candida gla	688	34	55.7	118	2	Q9C921_ARATH	Q9C921_arabidopsis

689	34	55.7	121	2	Q9NM9_PLAFA	Q9nm9 plasmodium	762	34	55.7	411	2	Q40N13_DESAC	Q40n13 desulfuromo
690	34	55.7	132	2	Q8TUB5_METAC	Q8tub5 methanosarc	763	34	55.7	423	2	Q3SQ87_NITWN	Q3sq87 nitrobacter
691	34	55.7	136	2	Q9LZH8_ARATH	Q9lzh8 arabidopsis	764	34	55.7	430	2	Q9VW98_DROME	Q9vw98 drosophila
692	34	55.7	139	2	Q2RRA9_RHORA	Q2rra9 rhodospiril	765	34	55.7	431	2	Q21650_CABEL	Q21650 caenorhabdi
693	34	55.7	142	2	Q6Z102_ORYSA	Q6z102 oryza sativ	766	34	55.7	440	2	Q2V3J2_ARATH	Q2v3j2 arabidopsis
694	34	55.7	178	2	Q6S944_PLAFA	Q6s944 plasmodium	767	34	55.7	457	2	Q656C8_ORYSA	Q656c8 oryza sativ
695	34	55.7	178	2	Q46RG7_RALEU	Q46rg7 ralestonia e	768	34	55.7	457	2	Q2S1L8_9GAMM	Q2s1l8 hanella che
696	34	55.7	178	2	Q3YN09_BACTK	Q3yn09 bacillus th	769	34	55.7	465	2	Q9VGA7_DROME	Q9vga7 drosophila
697	34	55.7	194	2	Q4SD36_TETNG	Q4sd36 tetradodon n	770	34	55.7	469	2	Y1J1_CABEL	Y1j1 caenorhabdi
698	34	55.7	200	2	Q3VID3_CHILB	Q3vid3 pelodictyon	771	34	55.7	469	2	Q6ACK3_LEIXX	Q6ack3 leifsonia x
699	34	55.7	200	2	Q4C3J1_CROWT	Q4c3j1 crocospaer	772	34	55.7	477	2	Q418U6_GIBZE	Q418u6 gibberella
700	34	55.7	216	2	Q2P835_XANOR	Q2p835 xanthomonas	773	34	55.7	480	2	Q9LON4_STRCO	Q9lon4 streptomyc
701	34	55.7	224	2	Q5H5E0_XANOR	Q5h5e0 xanthomonas	774	34	55.7	485	2	SAHHI_ARATH	SAHHI arabidopsis
702	34	55.7	224	2	Q9HJVI_STRCO	Q9hjvi streptomyce	775	34	55.7	485	2	SAHH2_ARATH	SAHH2 arabidopsis
703	34	55.7	240	2	Q836J1_ENTFA	Q836j1 enterococcu	776	34	55.7	485	2	SAHH_MESCR	SAHH mesembryant
704	34	55.7	243	2	Q3CEA5_SCLOR	Q3cea5 alkaliphilu	777	34	55.7	485	2	Q5D6C3_ARATH	Q5d6c3 arabidopsis
705	34	55.7	256	2	Q3F2H0_9BURK	Q3f2h0 burkholderi	778	34	55.7	485	2	Q5D6C4_ARATH	Q5d6c4 arabidopsis
706	34	55.7	260	2	Q450T1_9BURK	Q450t1 burkholderi	779	34	55.7	488	2	Q4RJVO_TETNG	Q4rjvo tetradodon n
707	34	55.7	260	2	Q4B8Z2_BURVI	Q4b8z2 burkholderi	780	34	55.7	488	2	Q7PNT2_ANOGA	Q7pnt2 anophelies g
708	34	55.7	260	2	Q4C3I9_CROWT	Q4c3i9 crocospaer	781	34	55.7	489	2	Q5YVSI_NOCFA	Q5yvsi nocardia fa
709	34	55.7	260	2	Q4LNA7_9BURK	Q4lna7 burkholderi	782	34	55.7	512	2	MBI3_YEAST	Q9zvw7 saccharomyc
710	34	55.7	260	2	Q39F58_BURS3	Q39f58 burkholderi	783	34	55.7	517	1	Q6BNE9_DEBHA	Q6bne9 debaromyce
711	34	55.7	260	2	Q62JD9_BURMA	Q62jd9 burkholderi	784	34	55.7	527	2	Q4RPR4_TETNG	Q4rpr4 tetradodon n
712	34	55.7	262	2	Q63T27_BURPS	Q63t27 burkholderi	785	34	55.7	533	2	Q7ZWU9_XENLA	Q7zwu9 xenopus lae
713	34	55.7	262	2	Q86276_ROTBS	Q86276 bovine rota	786	34	55.7	533	2	Q6GPB8_XENLA	Q6gpb8 xenopus lae
714	34	55.7	263	2	Q37JN9_RHOPA	Q37jn9 rhodopseudo	787	34	55.7	534	2	Q6P886_XENTR	Q6p886 xenopus tro
715	34	55.7	266	2	Q8LWZ9_9PHAE	Q8lwz9 laminaria d	788	34	55.7	534	2	Q57VH9_9TRYP	Q57vh9 trypanosoma
716	34	55.7	271	2	Q36291_BIJOU	Q36291 bijou bridg	789	34	55.7	545	2	Q51D85_ENTHI	Q51d85 entamoeba h
717	34	55.7	271	2	Q6AHQ6_CABEL	Q6ahq6 caenorhabdi	790	34	55.7	563	2	Q43KQ2_SOLUS	Q43kq2 solibacter
718	34	55.7	274	2	Q4ZCK8_9CAUD	Q4zck8 bacterioph	791	34	55.7	589	2	Q62547_LOLPE	Q62547 loligo peal
719	34	55.7	274	2	Q4ZCS6_9CAUD	Q4zcs6 bacterioph	792	34	55.7	591	2	APAFM_HUMAN	P43652 homo sapien
720	34	55.7	274	2	Q4ZD02_9CAUD	Q4zd02 bacterioph	793	34	55.7	599	2	Q32ME3_HUMAN	Q32mr3 streptomyc
721	34	55.7	274	2	Q8SDP2_9CAUD	Q8sdp2 staphylococ	794	34	55.7	602	2	Q9L1L2_STRCO	Q9l1l2 streptomyc
722	34	55.7	274	2	Q5HIY5_STAAC	Q5hiy5 staphylococ	795	34	55.7	602	2	Q7N4M8_PHOLL	Q7n4m8 photorhabdu
723	34	55.7	274	2	Q6GAK1_STAAS	Q6gak1 staphylococ	796	34	55.7	603	2	Q6FLM5_MESFL	Q6flm5 mesoplasma
724	34	55.7	274	2	Q6GGK5_STAAR	Q6ggk5 staphylococ	797	34	55.7	632	2	Q1QRI1_RAT	Q1qr1 rattus norv
725	34	55.7	274	2	Q8NWK9_STAAM	Q8nwk9 staphylococ	798	34	55.7	633	1	Q4Z362_PLABE	Q4z362 plasmodium
726	34	55.7	274	2	APAH_HAIN	P44751 haemophilus	799	34	55.7	651	2	Q7S668_NEUCR	Q7s668 neurospora
727	34	55.7	275	1	Q81EA3_BACCR	Q81ea3 bacillus ce	800	34	55.7	656	2	Q7FAF6_ORYSA	Q7faf6 oryza sativ
728	34	55.7	279	2	Q8EK17_SHEON	Q8ek17 shewanella	801	34	55.7	661	2	Q2PYR3_GVXN	Q2pyr3 xestia c-ni
729	34	55.7	279	2	Q2SWY3_BURTH	Q2swy3 burkholderi	802	34	55.7	661	2	Q41IW3_GIBZEL	Q41iw3 gibberella
730	34	55.7	281	2	Q3JR44_BURPL	Q3jr44 burkholderi	803	34	55.7	665	2	Q44062_AERHY	Q44062 aeromonas h
731	34	55.7	281	2	Q8AY27_YENLA	Q8ay27 xenopus lae	804	34	55.7	676	2	Q4UH19_THEAN	Q4uh19 theileria a
732	34	55.7	281	2	Q8GHW7_LACLC	Q8ghw7 lactococcus	805	34	55.7	677	2	Q26820_9TRYP	Q26820 trypanosoma
733	34	55.7	289	2	Q9CJ12_LACLA	Q9cj12 lactococcus	806	34	55.7	679	2	Q389A6_9TRYP	Q389a6 trypanosoma
734	34	55.7	289	2	Q9GTL4_DROSI	Q9gtl4 drosophila	807	34	55.7	682	2	Q9STF4_ARATH	Q9stf4 arabidopsis
735	34	55.7	292	2	DAPA_STRMU	Q8dues streptococc	808	34	55.7	712	2	Q7X8B0_ORYSA	Q7x8b0 oryza sativ
736	34	55.7	293	1	Q3KAT2_PSEPF	Q3kat2 pseudomonas	809	34	55.7	724	2	Q5OQN7_ENTHI	Q5oqn7 entamoeba h
737	34	55.7	294	2	Q2NXL2_SODGL	Q2nxl2 sodalis glo	810	34	55.7	729	2	Q5R1L2_IDILO	Q5r1l2 idiomarina
738	34	55.7	305	2	Q3NMV2_9GAMM	Q3nmv2 shewanella	811	34	55.7	733	2	Q6P7G8_XENLA	Q6p7g8 xenopus lae
739	34	55.7	307	2	Q4UQH2_XANC8	Q4uqh2 xanthomonas	812	34	55.7	736	1	KUP2_BIFLO	Q8g7q3 bifidobacte
740	34	55.7	308	2	Q8PCX8_XANCP	Q8pcx8 xanthomonas	813	34	55.7	751	1	SVY_NANEQ	Q74nf3 nanoarchaeu
741	34	55.7	308	2	Q75GK2_ORYSA	Q75gk2 oryza sativ	814	34	55.7	767	2	Q623C4_CABER	Q623c4 caenorhabdi
742	34	55.7	311	2	Q3VPPT_9CHLB	Q3vppt pelodictyon	815	34	55.7	770	1	YRN9_CABEL	Q09609 caenorhabdi
743	34	55.7	312	2	Q7P5H5_FUSNV	Q7p5h5 fusobacteri	816	34	55.7	777	2	Q2SJR4_9GAMM	Q2sjr4 habella che
744	34	55.7	320	2	Q8RHX5_FUSNV	Q8rhx5 fusobacteri	817	34	55.7	819	2	Q4WJ25_ASPFU	Q4wj25 aspergillus
745	34	55.7	320	2	YKYZ_BACSU	Q31690 bacillus su	818	34	55.7	825	2	Q4RT16_TETNG	Q4rt16 tetradodon n
746	34	55.7	321	1	Q85MB1_9FUNG	Q85mb1 monoblephar	819	34	55.7	868	2	Q3CW15_ALTAT	Q3cw15 pseudoalter
747	34	55.7	327	2	YSX3_CABEL	Q10022 caenorhabdi	820	34	55.7	912	2	Q9BDV6_ERIEU	Q9bdv6 erinaceus e
748	34	55.7	341	1	Q6AHQ7_CABEL	Q6ahq7 caenorhabdi	821	34	55.7	914	2	Q40PR3_DESAC	Q40pr3 desulfuromo
749	34	55.7	342	2	Q4HU81_GIBZE	Q4hu81 gibberella	822	34	55.7	928	2	Q2QVF3_ORYSA	Q2qvf3 oryza sativ
750	34	55.7	354	2	Q5CIE4_CRYHO	Q5cie4 cryptospori	823	34	55.7	932	2	Q2QVF5_SOLUT	Q2qvf5 solanum tub
751	34	55.7	361	2	Q8IQ21_DROME	Q8iq21 drosophila	824	34	55.7	937	2	Q2QVF5_SOLUT	Q2qvf5 solanum tub
752	34	55.7	362	2	Q51AH0_ENTHI	Q51ah0 entamoeba h	825	34	55.7	938	2	Q4XGY0_DICDI	Q4xgy0 dictyostell
753	34	55.7	363	2	Q7MB10_PHOLL	Q7mb10 photorhabdu	826	34	55.7	941	2	Q8T5Q4_MANSE	Q8t5q4 manduca sex
754	34	55.7	363	2	Q2V3J3_ARATH	Q2v3j3 arabidopsis	827	34	55.7	947	2	Q6C6Q8_YARLI	Q6c6q8 yarowia li
755	34	55.7	364	2	Q3FLV8_ARATH	Q3flv8 arabidopsis	828	34	55.7	952	2	Q4UFW1_THEAN	Q4ufw1 theileria a
756	34	55.7	373	2	Q36758_YEAST	Q36758 saccharomyc	829	34	55.7	993	2	Q9C2U8_GIBFU	Q9c2u8 gibberella
757	34	55.7	374	2	Q73XF7_MYCPA	Q73xf7 mycobacteri	830	34	55.7	1041	2	Q05925_NEIME	Q05925 neisseria m
758	34	55.7	382	2	Q6U7X9_CRIPE	Q6u7x9 crinipellis	831	34	55.7	1048	2	Q4P6A3_USTMA	Q4p6a3 ustilago ma
759	34	55.7	385	2	Q753U2_ASHGO	Q753u2 ashbya goss	832	34	55.7	1067	2	Q94046_CABEL	Q94046 caenorhabdi
760	34	55.7	400	2	Q4P4M6_USTMA	Q4p4m6 ustilago ma	833	34	55.7	1074	2		
761	34	55.7	402	2			834	34	55.7				

835	34	55.7	1076	2	Q9U370_CABEL	Q9U370 caenorhabdi	908	185	1	NUSG_HAEIN	P43916 haemophilus
836	34	55.7	1130	2	Q50WS9_ENTHI	Q50ws9 entamoeba h	909	185	2	Q3AW29_SYNS9	Q3aw29 synechococc
837	34	55.7	1159	2	Q4XUS6_PLACH	Q4xus6 plasmodium	910	185	2	Q4QMK2_HAE18	Q4qmk2 haemophilus
838	34	55.7	1362	2	Q55CB6_DICDI	Q55cb6 dictyosteli	911	187	2	Q67282_AQUAE	Q67282 aquifex aeo
839	34	55.7	1437	1	Q5GLM_BUNGE	P12430 bunyavirus	912	188	2	Q7V4W1_PROMM	Q7v4w1 prochloroco
840	34	55.7	1588	2	Q4QANS_LEIMA	Q4qans leishmania	913	190	2	Q7UE69_RHOBA	Q7ue69 rhodopirell
841	34	55.7	1647	2	Q54F23_DICDI	Q54f23 dictyosteli	914	191	2	Q3AMY2_SYNSC	Q3amy2 synechococc
842	34	55.7	1690	2	Q60YTA_CABER	Q60yta caenorhabdi	915	193	2	Q4FAC8_9SALA	Q4fac8 eurycea gut
843	34	55.7	1901	2	Q9DHH8_YLDV	Q9dhh8 yaba-like d	916	199	2	Q6CUX4_KLULA	Q6cux4 kluyveromyc
844	34	55.7	2135	2	Q61077_PLAFA	Q61077 plasmodium	917	206	2	Q7VIME_PROMP	Q7vime prochloroco
845	34	55.7	2539	2	Q5FIP8_LACAC	Q5fip8 lactobacill	918	211	2	Q4GTD4_9TRIP	Q4gtd4 trypanosoma
846	34	55.7	3133	2	Q6BS09_DEBHA	Q6bs09 debaryomyc	919	211	2	Q8Q209_9HIV1	Q8q209 human immun
847	33.5	54.9	105	2	Q9Q6M9_9DELA	Q9q6m9 human t-lym	920	212	2	Q8GEP8_9BACT	Q8gep8 uncultured
848	33.5	54.9	255	2	Q9D5S6_MOUSE	Q9d5s6 mus musculu	921	216	2	Q3HY53_9BACT	Q3hy53 uncultured
849	33.5	54.9	437	2	Q7RM38_PLAYO	Q7rm38 plasmodium	922	216	2	Q49CMA_9HIV1	Q49cm4 human immun
850	33.5	54.9	470	2	Q3TDL1_MOUSE	Q3tdl1 mus musculu	923	224	1	VV_MUMPE1	P60167 mumps virus
851	33.5	54.9	504	2	Q9BDM1_PONPY	Q9bdm1 pongo pygma	924	224	1	VV_MUMPE	P30927 mumps virus
852	33.5	54.9	680	1	Y964_METUA	Q58374 methanococc	925	224	1	VV_MUMPM	P30928 mumps virus
853	33.5	54.9	702	2	Q8SY79_DROME	Q8sy79 drosophila	926	224	1	VV_MUMPS	P33483 mumps virus
854	33.5	54.9	702	2	Q9VKV6_DROME	Q9kvk6 drosophila	927	224	2	Q6E4U2_9PARA	Q6e4u2 mumps virus
855	33.5	54.9	873	2	Q4CGH7_CLOTM	Q4cgh7 clostridium	928	224	2	Q6BC98_9PARA	Q6bc98 mumps virus
856	33.5	54.9	882	2	Q8MTD1_NASBA	Q8mtd1 maecigamoeb	929	224	2	Q77IS6_MUMPEJ	Q77is6 mumps virus
857	33.5	54.9	1002	2	Q74674_PNEJI	Q74674 pneumocysti	930	224	2	Q783W0_9PARA	Q783w0 mumps virus
858	33.5	54.9	1162	2	Q7P9W1_RICSI	Q7p9w1 rickettsia	931	224	2	Q83620_9PARA	Q83620 mumps virus
859	33.5	54.9	1162	2	Q9DGA5_RICCN	Q9dga5 rickettsia	932	224	2	Q8QV69_MUMPEJ	Q8qv69 mumps virus
860	33.5	54.9	1169	1	Y785_RICPR	O05975 rickettsia	933	224	2	Q91OG9_9PARA	Q91og9 mumps virus
861	33.5	54.9	1169	2	Q68W3_RICTY	Q68wv3 rickettsia	934	224	2	Q9DOA4_9PARA	Q9dqa4 mumps virus
862	33.5	54.9	1528	1	ZFYI16_MOUSE	Q8ou44 mus musculu	935	224	2	Q9J4L5_9PARA	Q9j4l5 mumps virus
863	33	54.1	59	2	Q1JIB7_BACCR	Q1jib7 bacillus ce	936	225	1	ACE1_YEAST	P15315 saccharomyc
864	33	54.1	59	2	Q81W13_BACAN	Q81w13 bacillus an	937	228	2	Q6U955_9CAUD	Q6u955 aeromonas p
865	33	54.1	69	2	Q7MES6_VIBVY	Q7mes6 vibrio vuln	938	229	2	Q8SML5_9CHLO	Q8sml5 dunaliella
866	33	54.1	86	1	SCR11_ARATH	P82630 arabidopsis	939	230	2	Q5E8Q9_VIBF1	Q5e8q9 vibrio fisc
867	33	54.1	103	2	Q9KAY6_BACHD	Q9kay6 bacillus ha	940	238	2	Q7NM03_GLOVI	Q7nm09 gloeobacter
868	33	54.1	109	2	Q4UDJ7_RICFE	Q4udj7 rickettsia	941	243	2	Q54ED0_DICDI	Q54ed0 dictyosteli
869	33	54.1	126	2	Q8T4Y7_PLACH	Q8t4y7 plasmodium	942	245	1	YAP5_YEAST	P40574 saccharomyc
870	33	54.1	127	2	Q5DBB4_SCHJA	Q5dbb4 schistosoma	943	245	1	Q4UD44_THEAN	Q4ud44 theileria a
871	33	54.1	132	2	Q4DI06_TRYCR	Q4di06 trypanosoma	944	246	2	Q4XJ25_PLACH	Q4xj25 plasmodium
872	33	54.1	132	2	Q4DU91_TRYCR	Q4du91 trypanosoma	945	249	2	Q6AL84_DESPS	Q6al84 desulfofale
873	33	54.1	136	2	Q8TS09_PLAFA	Q8ts09 plasmodium	946	251	2	Q3LTN9_9GEMI	Q3ltn9 cabbage lea
874	33	54.1	140	2	Q5AJG7_CANAL	Q5ajg7 candida alb	947	254	2	Q7PPX0_ANOGA	P10580 zea mays (m
875	33	54.1	140	2	Q6Q3Q4_SCORO	Q6q3q4 infectious	948	256	1	YMS4_MAIZE	Q9mj33 zea mays (m
876	33	54.1	140	2	Q6Q3Q5_SCORO	Q6q3q5 infectious	949	256	2	Q9MJC3_MAIZE	Q9mj33 zea mays (m
877	33	54.1	140	2	Q6Q3Q6_SCORO	Q6q3q6 infectious	950	266	2	Q4I2A4_GIBZE	Q4i2a4 gibberella
878	33	54.1	140	2	Q6Q3Q8_SCORO	Q6q3q8 infectious	951	269	2	Q5WTM9_LEGPL	Q5wtm9 legionella
879	33	54.1	140	2	Q6Q3R1_SCORO	Q6q3r1 infectious	952	275	2	Q3VSP5_PROAE	Q3vsp5 prosthecoch
880	33	54.1	140	2	Q6TN32_SCORO	Q6tn32 infectious	953	277	2	Q7P220_FUSNV	Q7p220 fusobacteri
881	33	54.1	140	2	Q6TN33_SCORO	Q6tn33 infectious	954	288	2	Q6C1C4_YARLI	Q6c1c4 yarrowia li
882	33	54.1	140	2	Q6TN35_SCORO	Q6tn35 infectious	955	289	2	Q5NHB0_FRATT	Q5nhb0 franciella
883	33	54.1	140	2	Q6TN36_SCORO	Q6tn36 infectious	956	290	2	Q5DDM7_SCHJA	Q5ddm7 schistosoma
884	33	54.1	140	2	Q6TN37_SCORO	Q6tn37 infectious	957	290	2	Q7MJF5_VIBVY	Q7mjf5 vibrio vuln
885	33	54.1	140	2	Q6TN38_SCORO	Q6tn38 infectious	958	290	2	Q82LN9_STRAW	Q82ln9 streptomyce
886	33	54.1	140	2	Q6TN39_SCORO	Q6tn39 infectious	959	290	2	Q9S1Q5_STRCO	Q9s1q5 streptomyce
887	33	54.1	144	2	Q6TN39_SCORO	Q6tn39 infectious	960	290	2	Q8K2N4_MOUSE	Q8k2n4 mus musculu
888	33	54.1	145	2	Q9XXT0_CAEEL	Q9xxt0 caenorhabdi	961	295	2	Q9H6Q3_HUMAN	Q9h6q3 homo sapien
889	33	54.1	145	2	Q7PT57_ANOGA	Q7pt57 anopheles g	962	300	2	Q8IEF4_PLAF7	Q8ief4 plasmodium
890	33	54.1	146	1	ANGI_COLGU	Q86ly5 colobus gule	963	300	2	Q5Z9F9_ORYSA	Q5z9f9 oryza sativ
891	33	54.1	149	2	Q7QTI6_GIALA	Q7qti6 giardia lam	964	300	2	Q8A974_BACTN	Q8a974 bacteroides
892	33	54.1	151	2	Q4HM34_CAMILA	Q4hm34 campylobact	965	300	2	Q8RG04_FUSNN	Q8rg04 fusobacteri
893	33	54.1	153	2	Q3VRN2_9SPHN	Q3vbn2 sphingopyxi	966	302	2	Q6BPK4_DEBHA	Q6bpk4 debaryomyc
894	33	54.1	154	2	Q3CAS9_9CLOT	Q3cas9 alkaliphilu	967	310	2	Q5EDW1_EMENI	Q5edw1 aspergillu
895	33	54.1	163	2	Q4ULI9_9SMEG	Q4uli9 punctigius p	968	313	2	Q65N94_BACLD	Q65n94 bacillus li
896	33	54.1	164	2	Q23462_ARABDPS	Q23462 arabidopsis	969	314	2	Q6MYT0_ASPPU	Q6myt0 aspergillu
897	33	54.1	165	2	Q8TBR1_DROME	Q8tbr1 drosophila	970	315	2	Q69BT6_EMENI	Q69bt6 emericella
898	33	54.1	172	1	IPYR_SYNEL	Q54y83 synechococc	971	316	2	Q2U2Y3_ASPOR	Q2u2y3 aspergillu
899	33	54.1	175	2	Q54VS3_DICDI	Q54vs3 dictyosteli	972	316	2	Q2WIS2_CLOPE	Q2wis2 clostridum
900	33	54.1	175	2	Q9M724_ARATH	Q9m724 arabidopsis	973	318	2	Q9PTX7_LAMRE	Q9ptx7 lampetra re
901	33	54.1	179	2	Q4LJX2_ARATH	Q4ljx2 arabidopsis	974	319	2	Q3CAQ8_9CLOT	Q3caq8 alkaliphilu
902	33	54.1	180	2	Q4YV59_PLABE	Q4yv59 plasmodium	975	321	2	Q9U5G0_EPTBU	Q9u5g0 eptaretus
903	33	54.1	180	2	Q7RRAS_PLAYO	Q7rras plasmodium	976	326	2	Q5E8X8_VIBF1	Q5e8x8 vibrio fisc
904	33	54.1	181	1	COTE_BACSU	P14016 bacillus su	977	330	1	YBFG_ECOLI	P37749 escherichia
905	33	54.1	183	2	Q4I4M8_GIBZE	Q4i4m8 gibberella	978	330	2	Q380N7_ANOGA	Q380n7 anopheles g
906	33	54.1	184	2	Q38ZK2_LACSS	Q38zk2 lactobacill	979	331	2	Q3NEA3_HUMAN	Q3nea3 homo sapien
907	33	54.1	184	2	Q7U9H8_SYNPX	Q7u9h8 synechococc	980	331	2	Q9U2B4_CABEL	Q9u2b4 caenorhabdi

981 33 54.1 331 2 Q483Q7\_COLP3 Q483q7 colwellia p  
 982 33 54.1 332 2 Q3QEL7\_9GAMM Q3qel7 shewanella  
 983 33 54.1 333 2 Q3IF68\_PSEHT Q3if68 pseudocalter  
 984 33 54.1 334 1 ADD1\_VIBPA Q87tf3 vibrio para  
 985 33 54.1 334 1 ADD1\_VIBPA Q9kni7 vibrio chol  
 986 33 54.1 336 2 Q977B2\_SULTO Q977b2 eulfolobus  
 987 33 54.1 344 2 Q6BWE5\_DEBHA Q6bwe5 debaryomyce  
 988 33 54.1 345 2 Q9CM15\_PASMU Q9cm15 pasteurella  
 989 33 54.1 347 2 Q49YJ9\_STAS1 Q49y9 staphylococ  
 990 33 54.1 349 2 Q7JNV6\_CAEEL Q7jnv6 caenorhabdi  
 991 33 54.1 349 2 Q17377\_CAEEL Q17377 caenorhabdi  
 992 33 54.1 349 2 Q34XY6\_9GAMM Q34xy6 alkalilimni  
 993 33 54.1 350 2 Q42928\_SCHFO Q42928 schizosacch  
 994 33 54.1 350 2 Q6LLR1\_PHOPR Q6llr1 photobacter  
 995 33 54.1 352 2 Q44BE2\_SOLDUS Q44be2 solibacter  
 996 33 54.1 352 2 Q6C4T3\_YARLI Q6c4t3 yarrowia li  
 997 33 54.1 357 2 Q41IM9\_METHU Q41im9 methanococ  
 998 33 54.1 357 2 Q7JNV7\_CAEEL Q7jnv7 caenorhabdi  
 999 33 54.1 357 2 Q9DF34\_BRARE Q9df34 brachydanio  
 1000 33 54.1 359 1 DPO4\_CLOPE Q9kx37 clostridium

## ALIGNMENTS

RESULT 1  
 HPSE HUMAN  
 ID HPSE\_HUMAN STANDARD; PRT; 543 AA.  
 AC Q9Y251; Q53GE5; Q9UL39;  
 DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
 DT 01-NOV-1999, sequence version 1.  
 DT 07-FEB-2006, entry version 27.  
 DE Heparanase precursor (EC 3.2.-.-) (Heparanase-1) (Hpal) (Endo-  
 DE glucuronidase) [Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa  
 DE subunit].  
 GN Name=HPSE; Synonyms=HEP, HPA, HPA1, HPR1, HPSE1, HSE1;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.  
 RC TISSUE=Placenta;  
 RX MEDLINE=99335379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;  
 RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C.,  
 RA Seddon A.P., Giorgio N.A., Bohlen P.;  
 RT "Cloning and functional expression of a human heparanase gene.";  
 RL Biochem. Biophys. Res. Commun. 261:183-187(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN  
 RP SEQUENCE OF 158-168; 326-337 AND 447-491.  
 RC TISSUE=Embryonic fibroblast;  
 RX MEDLINE=99337052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;  
 RA Toyoshima M., Nakajima M.;  
 RT "Human heparanase. Purification, characterization, cloning, and  
 RT expression.";  
 RL J. Biol. Chem. 274:24153-24160(1999).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.  
 RP PubMed=10395325; DOI=10.1038/10518;  
 RX Vlodavsky I., Friedmann Y., Elkin M., Aingorn H., Atzmon R.,  
 RX Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,  
 RA Spector L., Pecker I.;  
 RT "Mammalian heparanase: gene cloning, expression and function in tumor  
 RT progression and metastasis.";  
 RL Nat. Med. 5:793-802(1999).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE  
 RP OF 158-174; 263-272; 326-337; 433-436; 438-443; 466-468 AND 478-483.  
 RC TISSUE=Placenta;  
 RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;  
 RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,

Parish C.R.;  
 RT "Cloning of mammalian heparanase, an important enzyme in tumor  
 RT invasion and metastasis.";  
 RL Nat. Med. 5:803-809(1999).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE  
 RP SPECIFICITY.  
 RC TISSUE=Placenta;  
 RX MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;  
 RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;  
 RT "Heparanase expression in invasive trophoblasts and acute vascular  
 RT damage.";  
 RL Glycobiology 10:467-475(2000).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.  
 RP PubMed=11547900; DOI=10.1023/A:1011375624902;  
 RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O.,  
 RA Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S.,  
 RA Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M.,  
 RA Parekh R., Stubberfield C.;  
 RA "Biochemical characterization of the active heterodimer form of human  
 RT heparanase (Hpal) protein expressed in insect cells.";  
 RL Biochem. J. 373:423-435(2003).  
 RN [8]  
 RP NUCLEOTIDE SEQUENCE [MRNA].  
 RA Pinhal M.A., Semedo P.;  
 RT "Cloned heparanase from MCF-7 cells.";  
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.  
 RN [9]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
 RA Tanaka A., Yokoyama S.;  
 RT Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.  
 RN [10]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
 RC TISSUE=Pancras;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [11]  
 RP MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.  
 RX PubMed=11123890; DOI=10.1021/bi002080p;  
 RA Hulett M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,  
 RA Gready J.E., Parish C.R.;



DR EMBL; AF281160; AAF87301.2; -; mRNA.  
 DR InterPro; IPR005199; Glyco\_hydro\_79\_N.  
 DR Pfam; PF03662; Glyco\_hydro\_79n; 1.  
 KW Calcium; Glycoprotein; Hydrolase; Lysozyme; Magnesium; Membrane;  
 FT SIGNAL 1 37 By similarity.  
 FT CHAIN 38 111 Heparanase 8 kDa subunit (By similarity).  
 FT /FtId=PRO\_0000042256.  
 FT PROPEP 112 159 Linker peptide.  
 FT /FtId=PRO\_0000042257.  
 FT CHAIN 160 545 Heparanase 50 kDa subunit (By similarity).  
 FT /FtId=PRO\_0000042258.  
 FT REGION 160 164 Heparin/HS-binding (Potential).  
 FT REGION 272 282 Heparin/HS-binding (Potential).  
 FT ACT\_SITE 227 227 Proton donor (Potential).  
 FT ACT\_SITE 345 345 Nucleophile (Potential).  
 FT CARBOHYD 164 164 N-linked (GLCNAC...) (Potential).  
 FT CARBOHYD 219 219 N-linked (GLCNAC...) (Potential).  
 FT CARBOHYD 461 461 N-linked (GLCNAC...) (Potential).  
 SQ SEQUENCE 545 AA; 61077 MW; FAC4BDFD855B933 CRC64;

Query Match 80.3%; Score 49; DB 1; Length 545;  
 Best Local Similarity 80.0%; Pred.No. 4;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNPYK 10  
 |||||:||||  
 DB 439 CTNTGHPYK 448

RESULT 3  
 Q3R2P1 XYLFA  
 ID Q3R2P1 XYLFA PRELIMINARY; PRT; 567 AA.  
 AC Q3R2P1;  
 DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.  
 DT 25-OCT-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 3.  
 DE Aconitate hydratase (EC 4.2.1.3) (Fragment).  
 GN ORFNames=XfasoDRAFT\_1379;  
 OS Xylella fastidiosa Ann-1.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylella.  
 OX NCBI\_TaxID=155920;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Ann-1;  
 RG US DOE Joint Genome Institute (JGI-PGF);  
 RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,  
 RA Hammon N., Israni S., Pitluck S., Richardson P.,  
 RA "Sequencing of the draft genome and assembly of Xylella fastidiosa  
 RT Ann-1";  
 RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Ann-1;  
 RG US DOE Joint Genome Institute (JGI-ORNL);  
 RA Larimer F., Land M.;  
 RA "Annotation of the draft genome assembly of Xylella fastidiosa Ann-  
 RT 1";  
 RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Ann-1;  
 RG US DOE Joint Genome Institute (JGI-PGF);  
 RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,  
 RA Hammon N., Israni S., Pitluck S., Richardson P.;  
 RA Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.  
 RL [4]

CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; AAAM03000062; EAQ31487.1; -; Genomic DNA.  
 DR GO; GO:0003994; F:aconitate hydratase activity; IEA.  
 DR GO; GO:0016829; F:lyase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR000573; Aconitase\_C.  
 DR InterPro; IPR01030; Aconitase\_N.  
 DR Pfam; PF00330; Aconitase; 1.  
 DR Pfam; PF00694; Aconitase C; 1.  
 DR PRINTS; PR00415; ACONITASE.  
 DR ProDom; PD000511; Aconitase\_N; 1.  
 KW Lyase.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 567 AA; 61337 MW; 8F9E4B7AFCE7017E CRC64;

Query Match 70.5%; Score 43; DB 2; Length 567;  
 Best Local Similarity 87.5%; Pred.No. 51;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8  
 |||||  
 DB 104 CTNTSNP 111

RESULT 4  
 Q3R890 XYLFA  
 ID Q3R890 XYLFA PRELIMINARY; PRT; 608 AA.  
 AC Q3R890;  
 DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.  
 DT 25-OCT-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 3.  
 DE Aconitate hydratase, N-terminal.  
 GN ORFNames=XfasoDRAFT\_3065;  
 OS Xylella fastidiosa Ann-1.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylella.  
 OX NCBI\_TaxID=155920;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Ann-1;  
 RG US DOE Joint Genome Institute (JGI-PGF);  
 RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,  
 RA Hammon N., Israni S., Pitluck S., Richardson P.;  
 RA "Sequencing of the draft genome and assembly of Xylella fastidiosa  
 RT Ann-1";  
 RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Ann-1;  
 RG US DOE Joint Genome Institute (JGI-ORNL);  
 RA Larimer F., Land M.;  
 RA "Annotation of the draft genome assembly of Xylella fastidiosa Ann-  
 RT 1";  
 RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Ann-1;  
 RG US DOE Joint Genome Institute (JGI-PGF);  
 RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,  
 RA Hammon N., Israni S., Pitluck S., Richardson P.;  
 RA Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.  
 RL [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Ann-1;  
 RG US DOE Joint Genome Institute (JGI-PGF);  
 RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,  
 RA Hammon N., Israni S., Pitluck S., Richardson P.;  
 RA Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.  
 RL -!- CAUTION: The sequence shown here is derived from an



```
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AAAM03000011; EAO33483.1; -; Genomic DNA.
CC DR GO; GO:0016829; F:lyase activity; IEA.
CC DR GO; GO:0008152; P:metabolism; IEA.
CC DR InterPro; IPR001030; Aconitase_N.
CC DR Pfam; PF00330; Aconitase; 1.
CC DR PRINTS; PR00415; ACONITASE.
CC DR ProDom; PD000511; Aconitase_N; 1.
CC SQ SEQUENCE 608 AA; 64732 MW; FDIAB95AD0A45DD4 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 608;
Best Local Similarity 87.5%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 438 CTNTSNPR 445

RESULT 5
Q3QL63_9GAMM
ID Q3QL63_9GAMM PRELIMINARY; PRT; 861 AA.
AC Q3QL63;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, C-terminal:Aconitate hydratase, N-terminal.
GN ORFNames=SanadRAFT_1319;
OS Shewanella amazonensis SB28.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=326297;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SB28;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Izrani S., Pitluck S., Richardson P.;
RA "Sequencing of the draft genome and assembly of Shewanella amazonensis
RT SB28.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SB28;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RA "Annotation of the draft genome assembly of Shewanella amazonensis
RT SB28.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AAIN01000001; EAA40876.1; -; Genomic DNA.
CC DR GO; GO:0016836; F:hydro-lyase activity; IEA.
CC DR GO; GO:0005506; F:iron ion binding; IEA.
CC DR GO; GO:0016829; F:lyase activity; IEA.
CC DR InterPro; IPR012708; 2met_isocit_dhyd.
CC DR InterPro; IPR012084; Aco_LysF.
CC DR InterPro; IPR000573; Aconitase_C.
CC DR Pfam; PF00330; Aconitase; 1.
CC DR Pfam; PF00694; Aconitase_C; 1.
CC DR PIRSF; PIRSF001417; LysF; 1.
```

```
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhyd; 1.
SQ SEQUENCE 861 AA; 94219 MW; 47750044C9B194A CRC64;

Query Match 70.5%; Score 43; DB 2; Length 861;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 406 CTNTSNPR 413

RESULT 6
Q33T54_9GAMM
ID Q33T54_9GAMM PRELIMINARY; PRT; 862 AA.
AC Q33T54;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, C-terminal:Aconitate hydratase, N-terminal.
GN ORFNames=ShewDRAFT_0853;
OS Shewanella sp. PV-4.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=323850;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PV-4;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Izrani S., Pitluck S., Richardson P.;
RA "Sequencing of the draft genome and assembly of Shewanella sp. PV-4.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PV-4;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RA "Annotation of the draft genome assembly of Shewanella sp. PV-4.";
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AALS01000019; EAP02293.1; -; Genomic DNA.
CC DR GO; GO:0016836; F:hydro-lyase activity; IEA.
CC DR GO; GO:0005506; F:iron ion binding; IEA.
CC DR GO; GO:0016829; F:lyase activity; IEA.
CC DR GO; GO:0008152; P:metabolism; IEA.
CC SQ SEQUENCE 862 AA; 94331 MW; 131DD5B7AE952023 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 862;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 406 CTNTSNPR 413

RESULT 7
Q4ZUP2_PSEU2
ID Q4ZUP2_PSEU2 PRELIMINARY; PRT; 862 AA.
AC Q4ZUP2;
DT 07-JUN-2005, integrated into UniProtKB/TrEMBL.
DT 07-JUN-2005, sequence version 1.
DT 07-FEB-2006, entry version 6.
DE Aconitate hydratase, C-terminal:Aconitate hydratase, N-terminal.
```



```

GN OrderedLocusNames=Psy_2087;
OS Pseudomonas syringae pv. syringae (strain B728a).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=205918;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=16043691; DOI=10.1073/pnas.0504930102;
RA Feil H., Feil W.S., Chain P., Larimer F., Dibartolo G., Copeland A.,
RA Lykidis A., Trong S., Nolan M., Golsman E., Thiel J., Malfatti S.,
RA Loper J.E., Lapidus A., Dettler J.C., Land M., Richardson P.M.,
RA Kyrpides N.C., Ivanova N., Lindov S.E.;
RA "Comparison of the complete genome sequences of Pseudomonas syringae
RT pv. syringae B728a and pv. tomato DC3000."
RL Proc. Natl. Acad. Sci. U.S.A. 102:11064-11069(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; CP000075; AAY37130.1; -; Genomic_DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dHYD.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dHY; 1.
KW Complete proteome.
SQ SEQUENCE 862 AA; 93915 MW; C89D1AD262127EBA CRC64;

Query Match 70.5%; Score 43; DB 2; Length 862;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 406 CTNTSNPR 413
|||||

RESULT 8
Q88KF4_PSEPK PRELIMINARY; PRT; 862 AA.
AC Q88KF4;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
DT 07-FEB-2006, entry version 11.
DE Aconitate hydratase, putative.
GN OrderedLocusNames=PP2336; ORFNames=PP_2336;
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX MEDLINE=22423060; PubMed=12534463;
RX DOI=10.1046/j.1462-2920.2002.00366.x;
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L.M., Beanan M.J., DeBoy R.T., Daugherty S.C., Kolonay J.F.,
RA Madupu R., Nelson W.C., White O., Peterson J.D., Khouri H.M.,
RA Hance I., Chris Lee P., Holtzapfel E.K., Scanlan D., Tran K.,
RA Moazzaz A., Uitterback T.R., Rizzo M., Lee K., Kosack D., Moestl D.,
RA Wedler H., Lauber J., Stjepandic D., Hohseisel J., Straetz M., Heim S.,
RA Kiewitz C., Eisen J.A., Timmis K.N., Duesterhoeft A., Tuemmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
metabolically versatile Pseudomonas putida KT2440."

```

```

RL Environ. Microbiol. 4:799-808(2002).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AE015451; AAN67949.1; -; Genomic_DNA.
DR TIGR; PP2336; -;
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dHYD.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR001030; Aconitase_C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dHY; 1.
KW Complete proteome.
SQ SEQUENCE 862 AA; 93962 MW; BCB84EA2AE8624B6 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 862;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 406 CTNTSNPR 413
|||||

RESULT 9
Q2X9A9_PSEPU PRELIMINARY; PRT; 863 AA.
AC Q2X9A9;
DT 10-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 10-JAN-2006, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, N-terminal.
GN ORFNames=PputDRAFT_0339;
OS Pseudomonas putida Fl.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=351746;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Fl;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hamon N., Israni S., Pittluck S., Richardson P., Richardson P.;
RT "Sequencing of the draft genome and assembly of Pseudomonas putida
RT Fl."
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Fl;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Pseudomonas putida Fl.";
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AALM01000071; EAP48361.1; -; Genomic_DNA.
SQ SEQUENCE 863 AA; 94162 MW; 82F2F8549795839C CRC64;

Query Match 70.5%; Score 43; DB 2; Length 863;

```

```
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
DB 407 CTNTSNR 414

RESULT 10
Q48JZ8_PSE14 PRELIMINARY; PRT; 863 AA.
AC Q48JZ8;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 1.
DT 13-SEP-2005, sequence version 5.
DE 2-methylisocitrate dehydratase, Fe/S-dependent (EC 4.2.1.99).
GN Name-acnd; OrderedLocusNames=PSPPH_2058;
OS Pseudomonas syringae pv. phaseolicola (strain 1448A / Race 6).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=264730;
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=16159782; DOI=10.1128/JB.187.18.6488-6498.2005;
RA Joardar V., Lindeberg M., Jackson R., Selengut J., Dodson R.,
RA Brinkac L.M., Daugherty S.C., DeBoy R.T., Durkin A.S.,
RA Gwinn Gligio M., Madupu R., Nelson W.C., Rosovitz M.J., Sullivan S.A.,
RA Crabtree J., Creasy T., Davidsen T.M., Haft D.H., Zafar N., Zhou L.,
RA Halpin R., Holley T., Khouri H.M., Feldblyum T.V., White O.,
RA Fraser C.M., Chatterjee A.K., Cartinhour S., Schneider D.,
RA Mansfield J., Collmer A., Buell R.;
RT "Whole-genome sequence analysis of Pseudomonas syringae pv.
RT phaseolicola 1448A reveals divergence among pathogens in genes
RT involved in virulence and transposition.";
RL J. Bacteriol. 187:6488-6498(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
EMBL: CP000058; AA233624.1; -; Genomic DNA.
DR GO: 0047456; F:2-methylisocitrate dehydratase activity; IEA.
DR GO: 0005506; F:iron ion binding; IEA.
DR GO: 00016829; F:lyase activity; IEA.
DR GO: 0008152; P:metabolism; IEA.
DR InterPro: IPR012708; 2met_isocit_dhyd.
DR InterPro: IPR012084; Aco_LysF.
DR InterPro: IPR001030; Aconitase_N.
DR Pfam: PF00330; Aconitase; 1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PRINTS: PR001417; LysF; 1.
DR PRINTS: PR00415; ACONITASE.
DR ProDom: PD000511; Aconitase_N; 1.
DR TIGRFAMs: TIGR02333; 2met_isocit_dhy; 1.
KW Complete proteome; Lyase.
SQ SEQUENCE 863 AA; 94098 MW; 18D906940783C488 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 863;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
DB 407 CTNTSNR 414

RESULT 11
Q2P7A3_XANOR PRELIMINARY; PRT; 863 AA.
AC Q2P7A3;
DT 07-FEB-2006, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, sequence version 1.
DT 07-FEB-2006, entry version 1.
DE Aconitase hydratase 1.
SQ SEQUENCE 863 AA; 94791 MW; ACP4454F1C74B97B CRC64;
```

```
GN Name=XO00819;
OS Xanthomonas oryzae pv. oryzae MAFF 311018.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=342109;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=MAFF 311018;
RC Ochiai H., Inoue Y., Takeya M., Sasaki A., Kaku H.;
RA "Genome sequence of Xanthomonas oryzae pv. oryzae suggests
RT contribution of large numbers of effector genes and insertion
RT sequences to its race diversity.";
RL Jpn. Agric. Res. Q. 39:275-287(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
EMBL: AP008229; BAE67574.1; -; Genomic DNA.
DR SEQUENCE 863 AA; 93934 MW; 93D16B0BE4CAF67B CRC64;

Query Match 70.5%; Score 43; DB 2; Length 863;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
DB 406 CTNTSNR 413

RESULT 12
Q87P72_VIBPA PRELIMINARY; PRT; 863 AA.
AC Q87P72;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
DT 07-FEB-2006, entry version 15.
DE Aconitase hydratase 1.
GN OrderedLocusNames=VPI646;
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739; DOI=10.1016/S0140-6736(03)12659-1;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Naito M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae.";
RL Lancet 361:743-749(2003).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
EMBL: BA000031; BACS9909.1; -; Genomic DNA.
DR GO: 0016836; F:hydro-lyase activity; IEA.
DR GO: 0005506; F:iron ion binding; IEA.
DR GO: 00016829; F:lyase activity; IEA.
DR GO: 0008152; P:metabolism; IEA.
DR InterPro: IPR012708; 2met_isocit_dhyd.
DR InterPro: IPR012084; Aco_LysF.
DR InterPro: IPR000573; Aconitase_C.
DR InterPro: IPR001030; Aconitase_N.
DR Pfam: PF00330; Aconitase; 1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PRINTS: PR001417; LysF; 1.
DR PRINTS: PR00415; ACONITASE.
DR ProDom: PD000511; Aconitase_N; 1.
DR TIGRFAMs: TIGR02333; 2met_isocit_dhy; 1.
KW Complete proteome.
SQ SEQUENCE 863 AA; 94791 MW; ACP4454F1C74B97B CRC64;
```

Query Match 70.5%; Score 43; DB 2; Length 863;  
 Best Local Similarity 87.5%; Pred. No. 81;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CTNTDNP 8  
 Db 410 CTNTSNP 417

RESULT 13  
 Q883RI\_PSES  
 ID Q883RI\_PSES PRELIMINARY; PRT; 863 AA.  
 AC Q883RI;  
 DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.  
 DT 01-JUN-2003, sequence version 1.  
 DT 07-FEB-2006, entry version 13.  
 DE Aconitase family protein.  
 DE Aconitase family protein.  
 OS Pseudomonas syringae pv. tomato.  
 GN OrderedLocusNames=PSPTO2289; ORFNames=PSPTO\_2289;  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OC NCBI\_TaxID=323;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=DC3000;  
 RX MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;  
 RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,  
 RA Gwinn M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F.,  
 RA Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H.,  
 RA Nelson W.C., Davidson T.M., Zafar N., Zhou L., Liu J., Yuan Q.,  
 RA Khouri H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,  
 RA Uterback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,  
 RA Deng W.-L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K.,  
 RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,  
 RA Bender C.L., White O., Fraser C.M., Collier A.;  
 RT "The complete genome sequence of the Arabidopsis and tomato pathogen  
 Pseudomonas syringae pv. tomato DC3000.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186 (2003).  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 CC EMBL; AS016853; AA055803.1; -; Genomic\_DNA.  
 DR TIGR; PSPTO2289;  
 DR BioCyc; PSYR22283:PSPTO2289-MONOMER; -;  
 DR GO; GO:0016836; F:hydro-lyase activity; IEA.  
 DR GO; GO:0005506; F:iron ion binding; IEA.  
 DR GO; GO:0016829; F:lyase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR012708; 2met\_isocit\_dHVD.  
 DR InterPro; IPR012084; Aco\_LysF.  
 DR InterPro; IPR000573; Aconitase\_C.  
 DR InterPro; IPR001030; Aconitase\_N.  
 DR Pfam; PF00694; Aconitase; 1.  
 DR Pfam; PF00694; Aconitase; 1.  
 DR PIRSF; PIRSF001417; LysF; 1.  
 DR PRINTS; PD00415; ACONITASE.  
 DR ProDom; PD000511; Aconitase\_N; 1.  
 DR TIGRFAMs; TIGR02333; 2met\_isocit\_dHY; 1.  
 DR TIGRFAMs; TIGR02333; 2met\_isocit\_dHY; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 863 AA; 94052 MW; 6580FB5E2ACE13E0 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 863;  
 Best Local Similarity 87.5%; Pred. No. 81;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CTNTDNP 8  
 Db 406 CTNTSNP 413

RESULT 14

Q8PND4\_XANAC  
 ID Q8PND4\_XANAC PRELIMINARY; PRT; 863 AA.  
 AC Q8PND4;  
 DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.  
 DT 01-OCT-2002, sequence version 1.  
 DT 07-FEB-2006, entry version 13.  
 DE Aconitase hydratase 1.  
 DE Name=acnA;  
 GN Name=acnA;  
 OS Xanthomonas axonopodis pv. citri.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 OC NCBI\_TaxID=92829;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=306 / ATCC 13902 / XV 101;  
 RX MEDLINE=2202145; PubMed=12024217; DOI=10.1038/417459a;  
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A.,  
 RA Almeida N.F. Jr., Alves L.M.C., do Amaral A.M., Bertolini M.C.,  
 RA Camargo L.E.A., Camarotte G., Cannavan F., Cardozo J., Chambergo F.,  
 RA Ciapina L.P., Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R.,  
 RA El-Dorri H., Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Tezza R.I.D.,  
 RA Ferro M.I.T., Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
 RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,  
 RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
 RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA Setubal J.C., Kitajima J.P.;  
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
 host specificities.";  
 RL Nature 417:459-463 (2002).  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 CC EMBL; AE011742; AAM36011.1; -; Genomic DNA.  
 DR GO; GO:0016836; F:hydro-lyase activity; IEA.  
 DR GO; GO:0005506; F:iron ion binding; IEA.  
 DR GO; GO:0016829; F:lyase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR012708; 2met\_isocit\_dHVD.  
 DR InterPro; IPR012084; Aco\_LysF.  
 DR InterPro; IPR000573; Aconitase\_C.  
 DR InterPro; IPR001030; Aconitase\_N.  
 DR Pfam; PF00330; Aconitase; 1.  
 DR Pfam; PF00694; Aconitase; 1.  
 DR PIRSF; PIRSF001417; LysF; 1.  
 DR PRINTS; PD00415; ACONITASE.  
 DR ProDom; PD000511; Aconitase\_N; 1.  
 DR TIGRFAMs; TIGR02333; 2met\_isocit\_dHY; 1.  
 DR TIGRFAMs; TIGR02333; 2met\_isocit\_dHY; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 863 AA; 93708 MW; 7504D1917ED80F4E CRC64;

Query Match 70.5%; Score 43; DB 2; Length 863;  
 Best Local Similarity 87.5%; Pred. No. 81;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CTNTDNP 8  
 Db 406 CTNTSNP 413

RESULT 15  
 Q2T374\_BURTH  
 ID Q2T374\_BURTH PRELIMINARY; PRT; 864 AA.  
 AC Q2T374;  
 DT 24-JAN-2006, integrated into UniProtKB/TrEMBL.  
 DT 24-JAN-2006, sequence version 1.  
 DT 07-FEB-2006, entry version 2.  
 DE 2-methylisocitrate dehydratase, Fe/S-dependent (EC 4.2.1.99).

```
GN Name=acnD; ORFNames=BTH_I12187;
OS Burkholderia thailandensis E264;
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OX Burkholderiaceae; Burkholderia; pseudomallei group.
NCBI_TaxID=271848;
[1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=E264;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Palmer N., Dodson R.,
RA Hickey E.K., Winn M., Dougherty B., Fleischmann R.D., Richardson D.,
RA Peterson J., Krelavag A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van-Vugt R., Adams M.D., Gocayne J.D., Weidman J., Uterback T.,
RA Watthey L., McDonald L., Artiach P., Bowman C., Garland S., Fujii C.,
RA Cotton M.D., Horst K., Tomb J.-F., Roberts K., Hatch B., Smith H.O.,
RA Venter J.C.;
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; CP000085; ABC34711.1; -; Genomic DNA.
DR GO; GO:0016829; F:lyase activity; IEA.
KW Lyase.
SQ SEQUENCE 864 AA; 94223 MW; 4C12B4081417726 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 406 CTNTSNP 413

RESULT 16
Q3BWH4_XANC5 PRELIMINARY; PRT; 864 AA.
AC Q3BWH4;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 1.
DE Aconitate hydratase (EC 4.2.1.3).
GN Name=acnA; OrderedLocusNames=XCV1158;
OS Xanthomonas campestris pv. vesicatoria (strain 85-10).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
NCBI_TaxID=316273;
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP PubMed=16237009; DOI=10.1128/JB.187.21.7254-7266.2005;
RA Thieme F., Koebnik R., Bekel T., Berger C., Boch J., Buettner D.,
RA Chaldana C., Gaigalat L., Goemann A., Kay S., Kirchner O., Lanz C.,
RA Linke B., McHardy A.C., Meyer F., Mittenhuber G., Nies D.H.,
RA Niesbach-Kloeegen U., Patschewski T., Rueckert C., Rupp O.,
RA Schniesker S., Schuster S.C., Vorhoefer F.J., Weber E., Puehler A.,
RA Bonas U., Bartels D., Kaiser O.;
RT "Insights into genome plasticity and pathogenicity of the plant
RT pathogenic bacterium Xanthomonas campestris pv. vesicatoria revealed
RT by the complete genome sequence."
RL J. Bacteriol. 187:7254-7266 (2005).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AM039952; CAJ22789.1; -; Genomic DNA.
DR GO; GO:0003994; F:aconitate hydratase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
KW Complete proteome; Lyase.
SQ SEQUENCE 864 AA; 94066 MW; ADBAD8BA90C6BDE7 CRC64;

Name=acnD; ORFNames=BTH_I12187;
OS Burkholderia thailandensis E264;
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OX Burkholderiaceae; Burkholderia; pseudomallei group.
NCBI_TaxID=271848;
[1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=E264;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Palmer N., Dodson R.,
RA Hickey E.K., Winn M., Dougherty B., Fleischmann R.D., Richardson D.,
RA Peterson J., Krelavag A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van-Vugt R., Adams M.D., Gocayne J.D., Weidman J., Uterback T.,
RA Watthey L., McDonald L., Artiach P., Bowman C., Garland S., Fujii C.,
RA Cotton M.D., Horst K., Tomb J.-F., Roberts K., Hatch B., Smith H.O.,
RA Venter J.C.;
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; CP000085; ABC34711.1; -; Genomic DNA.
DR GO; GO:0016829; F:lyase activity; IEA.
KW Lyase.
SQ SEQUENCE 864 AA; 94223 MW; 4C12B4081417726 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 406 CTNTSNP 413

RESULT 17
Q454U1_9BURK PRELIMINARY; PRT; 864 AA.
AC Q454U1;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 2.
DE Aconitate hydratase, C-terminal:Aconitate hydratase, N-terminal.
OS ORFNames=BeenDRAFT_3590;
GN Burkholderia cenocepacia AU 1054.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; Burkholderia cenocepacia complex.
NCBI_TaxID=331271;
[1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN-AU 1054;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Izrani S., Pitluck S., Richardson P.,
RT "Sequencing of the draft genome assembly of Burkholderia cenocepacia
RT AU 1054."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN-AU 1054;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Burkholderia cenocepacia
RT AU 1054."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AAH101000012; EM12040.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met.isocit_dhyd.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD004511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met.isocit_dhy; 1.
SQ SEQUENCE 864 AA; 93860 MW; 60CFA11E2A0133C9 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 406 CTNTSNP 413

RESULT 18
Q4BTS4_BURVI
```

ID Q4BTS4\_BURVI PRELIMINARY; PRT; 864 AA.  
AC Q4BTS4;  
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.  
DT 13-SEP-2005, sequence version 1.  
DT 07-FEB-2006, entry version 2.  
DE Aconitate hydratase, C-terminal; Aconitate hydratase, N-terminal.  
GN ORFNames=Bcep1808DRAFT\_7478;  
OS Burkholderia vietnamiensis G4.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Burkholderiaceae; Burkholderia; Burkholderia cepacia complex.  
OX NCBI\_TaxID=269482;  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=G4;  
RC US DOE Joint Genome Institute (JGI-PGF);  
RG Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,  
RA Hammon N., Israni S., Pittluck S., Richardson P.;  
RA "Sequencing of the draft genome and assembly of Burkholderia  
RT vietnamiensis G4";  
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=G4;  
RG US DOE Joint Genome Institute (JGI-PGF);  
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,  
RA Hammon N., Israni S., Pittluck S., Richardson P.;  
RA "Sequencing of the draft genome and assembly of Burkholderia  
RT vietnamiensis G4";  
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=G4;  
RG US DOE Joint Genome Institute (JGI-PGF);  
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,  
RA Hammon N., Israni S., Pittluck S., Richardson P.;  
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
CC -! CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonDerivs License  
CC -----  
CC EMBL; AAHE02000001; EAM33209.1; -; Genomic DNA.  
DR GO; GO:0016836; F:hydro-lyase activity; IEA.  
DR GO; GO:0005506; F:iron ion binding; IEA.  
DR GO; GO:0016829; F:lyase activity; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR012084; 2met\_isocit\_dHYD.  
DR InterPro; IPR01030; Aconitase\_N.  
DR Pfam; PF00330; Aconitase; 1.  
DR Pfam; PF00694; Aconitase\_C; 1.  
DR PIRSF; PIRSF001417; LysF; 1.  
DR PRINTS; PR00415; ACONITASE.  
DR ProDom; PD000511; Aconitase\_N; 1.  
DR TIGRFAMs; TIGR02333; 2met\_isocit\_dHY; 1.  
SQ SEQUENCE 864 AA; 94035 MW; 93B5A360E67F0FB1 CRC64;  
  
Query Match 70.5%; Score 43; DB 2; Length 864;  
Best Local Similarity 87.5%; Pred. No. 81;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 CTNTDNP 8  
Db |||||  
406 CTNTSNRP 413  
  
RESULT 19  
Q4LQNS\_9BURK PRELIMINARY; PRT; 864 AA.  
ID Q4LQNS\_9BURK  
AC Q4LQNS;  
DT 02-AUG-2005, integrated into UniProtKB/TrEMBL.  
DT 02-AUG-2005, sequence version 1.

DT 07-FEB-2006, entry version 2.  
DE Aconitate hydratase, C-terminal; Aconitate hydratase, N-terminal.  
GN ORFNames=Bcen2424DRAFT\_3126;  
OS Burkholderia cenocepacia HI2424.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Burkholderiaceae; Burkholderia; Burkholderia cepacia complex.  
OX NCBI\_TaxID=331272;  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=HI2424;  
RC US DOE Joint Genome Institute (JGI-PGF);  
RG Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,  
RA Hammon N., Israni S., Pittluck S., Richardson P.;  
RA "Sequencing of the draft genome assembly of Burkholderia cenocepacia  
RT HI2424";  
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=HI2424;  
RG US DOE Joint Genome Institute (JGI-ORNL);  
RA Larimer F., Land M.;  
RA "Annotation of the draft genome assembly of Burkholderia cenocepacia  
RT HI2424";  
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
CC -! CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonDerivs License  
CC -----  
CC EMBL; AAHL01000027; EAM18383.1; -; Genomic DNA.  
DR GO; GO:0016836; F:hydro-lyase activity; IEA.  
DR GO; GO:0005506; F:iron ion binding; IEA.  
DR GO; GO:0016829; F:lyase activity; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR012708; 2met\_isocit\_dHYD.  
DR InterPro; IPR012084; Aco\_LysF.  
DR InterPro; IPR01030; Aconitase\_N.  
DR Pfam; PF00330; Aconitase; 1.  
DR Pfam; PF00694; Aconitase\_C; 1.  
DR PIRSF; PIRSF001417; LysF; 1.  
DR PRINTS; PR00415; ACONITASE.  
DR ProDom; PD000511; Aconitase\_N; 1.  
DR TIGRFAMs; TIGR02333; 2met\_isocit\_dHY; 1.  
SQ SEQUENCE 864 AA; 93860 MW; 60CF411E2A0133C9 CRC64;  
  
Query Match 70.5%; Score 43; DB 2; Length 864;  
Best Local Similarity 87.5%; Pred. No. 81;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 CTNTDNP 8  
Db |||||  
406 CTNTSNRP 413  
  
RESULT 20  
Q4URR6\_XANC8 PRELIMINARY; PRT; 864 AA.  
ID Q4URR6\_XANC8  
AC Q4URR6;  
DT 05-JUL-2005, integrated into UniProtKB/TrEMBL.  
DT 05-JUL-2005, sequence version 1.  
DT 07-FEB-2006, entry version 5.  
DE Aconitate hydratase 1.  
GN OrderedLocusNames=XC\_3213;  
OS Xanthomonas campestris pv. campestris (strain 8004).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.  
OX NCBI\_TaxID=314565;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RX PubMed=15899963; DOI=10.1101/gr.3378705;  
RA Qian W., Jia Y., Ren S.-X., He Y.-Q., Feng J.-X., Lu L.-F., Sun Q.,

```
RA Ying G., Tang D.-J., Tang H., Wu W., Hao P., Wang L., Jiang B.-L.,
RA Zeng S., Gu W.-Y., Lu G., Rong L., Tian Y., Yao Z., Fu G., Chen B.,
RA Fang R., Qiang B., Chen Z., Zhao G.-P., Tang J.-L., He C.,
RT "Comparative and functional genomic analyses of the pathogenicity of
RL phytopathogen Xanthomonas campestris pv. campestris.";
RL Genome Res. 15:757-767(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000050; AAY50257.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR01030; Aconitase N.
DR Pfam; PF00694; Aconitase C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR PRODOM; PD000511; Aconitase N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
KW Complete proteome.
SQ SEQUENCE 864 AA; 94059 MW; 8CC8BBBA4E1696BF CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 406 CTNTSNPR 413

RESULT 21
Q8VPS8_9BURK BURK
ID Q8VPS8_9BURK PRELIMINARY; PRT; 864 AA.
AC Q8VPS8;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2002, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE AcnM.
GN Name=acnM;
OS Burkholderia sacchari.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=159450;
RN [1]
RZ STRAIN=IPT101T;
RC MEDLINE=21633825; PubMed=11772636; DOI=10.1128/AEM.68.1.271-279.2002;
RA Bramer C.O., Silva L.F., Gomez J.G.C., Priefer H., Steinbuechel A.;
RT "Identification of the 2-Methylcitrate Pathway Involved in the
RT Catabolism of Propionate in the Polyhydroxyalkanoate-Producing Strain
RT Burkholderia sacchari IPT101(T) and Analysis of a Mutant Accumulating
RT a Copolyester with Higher 3-Hydroxyvalerate Content.";
RL Appl. Environ. Microbiol. 68:271-279(2002).
RN [2]
RZ STRAIN=IPT101T;
RC Braemer C.O., Steinbuechel A.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY033092; AAK52341.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.

us-10-645-659a-6.rup
```

```
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR000573; Aconitase C.
DR InterPro; IPR001030; Aconitase N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR PRODOM; PD000511; Aconitase N; 1. dhy; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
SQ SEQUENCE 864 AA; 94376 MW; A058CED3BC95C48C CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 406 CTNTSNPR 413

RESULT 22
Q39BAL_BURS3 BUR3
ID Q39BAL_BURS3 PRELIMINARY; PRT; 864 AA.
AC Q39BAL;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 22-NOV-2005, sequence version 1.
DT 21-FEB-2006, entry version 4.
DE Aconitate hydratase-like (EC 4.2.1.3).
GN OrderedLocusNames=Bcep18194_B0143;
OS Burkholderia sp. (strain 383) (Burkholderia cepacia (strain ATCC 17760
OS / NCIB 9086 / R18194)).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; Burkholderia cepacia complex.
OX NCBI_TaxID=269483;
RN [1]
RZ NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RG US DOE Joint Genome Institute;
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Chain P., Malfatti S., Shin M.,
RA Vergez L., Schmutz J., Larimer F., Land M., Kyrpides N., Lykidis A.,
RA Richardson P.;
RA Richardson P.;
RT "Complete sequence of chromosome 2 of Burkholderia sp. 383.";
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000152; ABB10260.1; -; Genomic DNA.
DR GO; GO:0003994; F:aconitate hydratase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
KW Complete proteome; Lyase.
SQ SEQUENCE 864 AA; 94002 MW; AD8FC6C259002FE4 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
DB 406 CTNTSNPR 413

RESULT 23
Q3JHQ8_BURP1 BURP1
ID Q3JHQ8_BURP1 PRELIMINARY; PRT; 864 AA.
AC Q3JHQ8;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 1.
DT 21-FEB-2006, entry version 4.
DE 2-methylisocitrate dehydratase, Fe/S-dependent (EC 4.2.1.99).
```

GN Name-acnd; OrderedLocusNames=BURPS1710b\_A1738;  
OS Burkholderia pseudomallei (strain 1710b).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Burkholderiaceae; Burkholderia; pseudomallei group.  
OX NCBI\_TaxID=320372;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RA Woods D.E., Nierman W.C.;  
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; CP000125; ABA53629.1; -; Genomic\_DNA.  
DR TIGR; BURPS1710b\_A1738; -;  
DR GO; GO:0047456; F:2-methylisocitrate dehydratase activity; IEA.  
DR GO; GO:0005506; F:iron ion binding; IEA.  
DR GO; GO:0016829; F:lyase activity; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR012708; 2met\_isocit\_dhyd.  
DR InterPro; IPR012084; Aco\_LysF.  
DR InterPro; IPR000573; Aconitase\_C.  
DR InterPro; IPR001030; Aconitase\_N.  
DR Pfam; PF00330; Aconitase; 1.  
DR Pfam; PF00694; Aconitase\_C; 1.  
DR PIRSF; PIRSF001417; LysF; 1.  
DR PRINTS; PR00415; ACONITASE.  
DR ProDom; PD000511; Aconitase\_N; 1.  
DR TIGRFAMs; TIGR02333; 2met\_isocit\_dhy; 1.  
KW Complete proteome; Lyase.  
SQ SEQUENCE 864 AA; 93924 MW; F8D1F356678CB2B5 CRC64;  
  
Query Match 70.5%; Score 43; DB 2; Length 864;  
Best Local Similarity 87.5%; Pred. No. 81;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 CTNTDNP 8  
Db 406 CTNTSNPR 413  
|||||  
RESULT 24  
Q3KFE7\_PSEPF PRELIMINARY; PRT; 864 AA.  
AC Q3KFE7;  
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.  
DT 08-NOV-2005, sequence version 1.  
DT 21-FEB-2006, entry version 4.  
DE Aconitase hydratase-like.  
GN OrderedLocusNames=Pfl\_1766;  
OS Pseudomonas fluorescens (strain pfo-1).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=205922;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RG US DOE Joint Genome Institute;  
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,  
RA Hammon N., Israni S., Pitluck S., Saunders E.H., Schmutz J.,  
RA Larimer F., Land M., Kyriakides N., Anderson I., Richardson P.;  
RT "Complete sequence of Pseudomonas fluorescens pfo-1";  
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; CP000094; ABA73509.1; -; Genomic\_DNA.  
DR GO; GO:0016836; F:hydro-lyase activity; IEA.  
DR GO; GO:0005506; F:iron ion binding; IEA.  
DR GO; GO:0016829; F:lyase activity; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR012708; 2met\_isocit\_dhyd.  
DR InterPro; IPR012084; Aco\_LysF.

DR InterPro; IPR000573; Aconitase\_C.  
DR InterPro; IPR001030; Aconitase\_N.  
DR Pfam; PF00330; Aconitase; 1.  
DR Pfam; PF00694; Aconitase\_C; 1.  
DR PIRSF; PIRSF001417; LysF; 1.  
DR PRINTS; PR00415; ACONITASE.  
DR ProDom; PD000511; Aconitase\_N; 1.  
DR TIGRFAMs; TIGR02333; 2met\_isocit\_dhy; 1.  
KW Complete proteome.  
SQ SEQUENCE 864 AA; 94414 MW; 057777344276BD90 CRC64;  
  
Query Match 70.5%; Score 43; DB 2; Length 864;  
Best Local Similarity 87.5%; Pred. No. 81;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 CTNTDNP 8  
Db 406 CTNTSNPR 413  
|||||  
RESULT 25  
Q4KFKI\_PSEFS PRELIMINARY; PRT; 864 AA.  
AC Q4KFKI;  
DT 02-AUG-2005, integrated into UniProtKB/TrEMBL.  
DT 02-AUG-2005, sequence version 1.  
DT 07-FEB-2006, entry version 5.  
DE 2-methylisocitrate dehydratase, Fe/S-dependent (EC 4.2.1.99).  
GN Name=acnd; OrderedLocusNames=PFL1863;  
OS Pseudomonas fluorescens (strain Pf-5 / ATCC BAA-477).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=220664;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RX PubMed=15980861; DOI=10.1038/nbt1110;  
RA Paulsen I.T., Press C.M., Ravel J., Kobayashi D.Y., Myers G.S.A.,  
RA Mavrodi D.V., DeBoy R.T., Seshadri R., Ren Q., Madupu R., Dodson R.J.,  
RA Durkin A.S., Brinkac L.M., Daugherty S.C., Sullivan S.A.,  
RA Rosovitz M.J., Gwinn M.L., Zhou L., Schneider D.J., Cartinhour S.W.,  
RA Nelson W.C., Weidman J., Watkins K., Tran K., Khouri H., Pierson E.A.,  
RA Pierson L.S. III, Thomasow L.S., Loper J.E.;  
RT "Complete genome sequence of the plant commensal Pseudomonas  
fluorescens Pf-5.";  
RL Nat. Biotechnol. 23:873-878(2005).  
CC Copyrighted by the Uniprot Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; CP000076; AAY91151.1; -; Genomic\_DNA.  
DR GO; GO:0047456; F:2-methylisocitrate dehydratase activity; IEA.  
DR GO; GO:0005506; F:iron ion binding; IEA.  
DR GO; GO:0016829; F:lyase activity; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR012708; 2met\_isocit\_dhyd.  
DR InterPro; IPR012084; Aco\_LysF.  
DR InterPro; IPR001030; Aconitase\_N.  
DR Pfam; PF00330; Aconitase; 1.  
DR Pfam; PF00694; Aconitase\_C; 1.  
DR PIRSF; PIRSF001417; LysF; 1.  
DR PRINTS; PR00415; ACONITASE.  
DR ProDom; PD000511; Aconitase\_N; 1.  
DR TIGRFAMs; TIGR02333; 2met\_isocit\_dhy; 1.  
KW Complete proteome; Lyase.  
SQ SEQUENCE 864 AA; 94154 MW; C8F37A1B7BAB56D2 CRC64;  
  
Query Match 70.5%; Score 43; DB 2; Length 864;  
Best Local Similarity 87.5%; Pred. No. 81;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 CTNTDNP 8  
Db 406 CTNTSNPR 413  
|||||



```

RESULT 26
QS02UO IDILO PRELIMINARY; PRT; 864 AA.
ID QS02UO;
AC QS02UO;
DT 04-JAN-2005, integrated into UniProtKB/TrEMBL.
DT 04-JAN-2005, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE Aconitase A.
GN OrderedLocusNames=IL1425;
OS Idiomarina loihiensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Idiomarinaceae; Idiomarina.
OX NCBI_TaxID=135577;
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=L2-TR / DSM 15497 / ATCC BAA-735;
RC PubMed=15596722; DOI=10.1073/pnas.0407638102;
RX Hou S., Saw J.H., Lee K.S., Freitas T.A., Bellisle C., Kawarabayashi Y.,
RA Donachie S.P., Fikina A., Galperin M.Y., Koonin E.V., Makarova K.S.,
RA Omelchenko M.V., Sorokin A., Wolf Y.I., Li Q.X., Keum Y.S.,
RA Campbell S., Denery J., Aizawa S., Shibata S., Malahoff A., Alam M.;
RT "Genome sequence of the deep-sea gamma-proteobacterium Idiomarina
loihiensis reveals amino acid fermentation as a source of carbon and
energy."
RL Proc. Natl. Acad. Sci. U.S.A. 101:18036-18041(2004).
[1]
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
[1]
EMBL; AE017340; AAV82265.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR000573; Aconitase C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
KW Complete proteome.
SQ SEQUENCE 864 AA; 95071 MW; E3C54A825E79B6B CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 8
Db 406 CTNTSNP 413

RESULT 27
Q62A69 BURMA PRELIMINARY; PRT; 864 AA.
ID Q62A69 BURMA
AC Q62A69;
DT 25-OCT-2004, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2004, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE Aconitase hydratase (EC 4.2.1.3).
GN Name=acnA; Synonym=acn; OrderedLocusNames=BMAA1868;
OS Burkholderia mallei (Pseudomonas mallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=13373;
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=K96243;
RC PubMed=15377794; DOI=10.1073/pnas.0403302101;
RX Holden M.T.G., Titball R.W., Peacock S.J., Cerdano-Tarraga A.-M.,
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.L.,
RA Bentley S.D., Sebahia M., Thomson N.R., Bason N., Beacham I.R.,
RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,
RA Chillingworth T., Cronin A., Crosssett B., Davis P., Deshaizer D.,
RA Feltwell T., Fraser A., Hance Z., Hauser H., Holroyd S., Jagels K.,
RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,
RA Rabinowitz E., Rutherford K., Sanders M., Simmonds M.,
RA Songvilai S., Stevens K., Tumapa S., Vesatichavest M.,
RA Whitehead S., Yeats C., Barrell B.G., Oyston P.C.F., Parkhill J.;
RT "Genomic plasticity of the causative agent of melioidosis,
Burkholderia pseudomallei."
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).

```

```

RC STRAIN=ATCC 23344;
RX PubMed=15377793; DOI=10.1073/pnas.0403306101;
RA Nieman W.C., Deshaizer D., Kim H.S., Tettelin H., Nelson K.E.,
RA Feldblyum T.V., Ulrich R.L., Ronning C.M., Brinkac L.M., Dodson R.J.,
RA Daugherty S.C., Davidsen T.D., DeBoy R.T., Dimitrov G., Dodson R.J.,
RA Madupu A.S., Gwinn M.L., Haft D.H., Khouri H.M., Kolonay J.F.,
RA Sarria S., Selengut J., Shamblin C., Radune D., Romero C.M.,
RA Zafar N., Zhou L., Fraser C.M.;
RT "Structural flexibility in the Burkholderia mallei genome."
RL Proc. Natl. Acad. Sci. U.S.A. 101:14246-14251(2004).
[1]
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
[1]
EMBL; CP000011; AAU45605.1; -; Genomic DNA.
DR TIGR; BMAA1868; -;
DR GO; GO:0003994; F:aconitase hydratase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR000573; Aconitase C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
KW Complete proteome; Lyase.
SQ SEQUENCE 864 AA; 93962 MW; 39963B9C243B96C2 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 8
Db 406 CTNTSNP 413

RESULT 28
Q63NU1 BURPS PRELIMINARY; PRT; 864 AA.
ID Q63NU1 BURPS
AC Q63NU1;
DT 25-OCT-2004, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2004, sequence version 1.
DT 07-FEB-2006, entry version 9.
DE Aconitase hydratase 1 (EC 4.2.1.3).
GN Name=acnA; Synonym=acn; OrderedLocusNames=BPS0208;
OS Burkholderia pseudomallei (Pseudomonas pseudomallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; pseudomallei group.
OX NCBI_TaxID=28450;
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=K96243;
RC PubMed=15377794; DOI=10.1073/pnas.0403302101;
RX Holden M.T.G., Titball R.W., Peacock S.J., Cerdano-Tarraga A.-M.,
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.L.,
RA Bentley S.D., Sebahia M., Thomson N.R., Bason N., Beacham I.R.,
RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,
RA Chillingworth T., Cronin A., Crosssett B., Davis P., Deshaizer D.,
RA Feltwell T., Fraser A., Hance Z., Hauser H., Holroyd S., Jagels K.,
RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,
RA Rabinowitz E., Rutherford K., Sanders M., Simmonds M.,
RA Songvilai S., Stevens K., Tumapa S., Vesatichavest M.,
RA Whitehead S., Yeats C., Barrell B.G., Oyston P.C.F., Parkhill J.;
RT "Genomic plasticity of the causative agent of melioidosis,
Burkholderia pseudomallei."
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).

```

```
CC
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BX571966; CAH37653.1; -; Genomic DNA.
DR GO; GO:0003994; F:aconitate hydratase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; AcoLysF.
DR InterPro; IPR000573; Aconitase_C.
DR Pfam; PF00694; Aconitase_C; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
DR Complete proteome; Lyase.
KW SEQUENCE 864 AA; 93904 MW; 6C286C89554F6CB1 CRC64;
SQ
Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTNTDNP 8
DB 406 CTNTSNPR 413
RESULT 29
Q8PB75 XANCP
ID Q8PB75 XANCP PRELIMINARY; PRT; 864 AA.
AC Q8PB75;
DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2002, sequence version 1.
DT 07-FEB-2006, entry version 14.
DE Aconitate hydratase 1.
GN Name: achn;
OS Xanthomonas campestris pv. campestris.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]_TaxID=340;
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217; DOI=10.1038/417459a;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A.,
RA Almeida N.F., Jr., Alves L.M.C., do Amaral A.M., Bertolini M.C.,
RA Camargo L.E.A., Camarotte G., Cannavan F., Cardozo J., Chambergo F.,
RA Ciapina L.P., Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R.,
RA El-Dorry H., Faria J.B., Ferreira A.J.S., Ferreira R.C.C.,
RA Ferro M.I.T., Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
CC
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AE012199; AAM40332.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
```

```
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; AcoLysF.
DR InterPro; IPR000573; Aconitase_C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
DR Complete proteome.
KW SEQUENCE 864 AA; 94059 MW; 8C8BBBA4E1696BF CRC64;
SQ
Query Match 70.5%; Score 43; DB 2; Length 864;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTNTDNP 8
DB 406 CTNTSNPR 413
RESULT 30
Q3F693 9BURK
ID Q3F693 9BURK PRELIMINARY; PRT; 865 AA.
AC Q3F693;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, C-terminal; Aconitate hydratase, N-terminal.
GN ORFNames=BambDRAFT_1678;
OS Burkholderia ambifaria AMMD.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; Burkholderia cepacia complex.
OX NCBI_TaxID=339670;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AMMD;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Burkholderia ambifaria
RT AMMD."
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
RN [2]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AMMD;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Burkholderia ambifaria
RT AMMD."
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AJU101000010; EA044596.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; AcoLysF.
DR InterPro; IPR000573; Aconitase_C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
```

```

DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhyd; 1.
SQ SEQUENCE 865 AA; 94086 MW; A5BB8AD9789176A5 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 865;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
Db 407 CTNTSNPR 414

RESULT 31
Q3QY1_RALME
ID Q3QY1_RALME PRELIMINARY; PRT; 865 AA.
AC Q3QY1;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, C-terminal; Aconitate hydratase, N-terminal.
GN ORFNames=RmetDRAFT_4095;
OS Ralstonia metallidurans (strain CH34).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Cupriavidus.
OX NCBI_TaxID=266264;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CH34;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Ralstonia
RT metallidurans CH34.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
RN [2];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CH34;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Ralstonia metallidurans
RT CH34.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
RN [3];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CH34;
RG US DOE Joint Genome Institute;
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Richardson P.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
EMBL; AAA10300006; EAN50810.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR00573; Aconitase C.
DR EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
EMBL; AL646076; CAD17271.1; -; Genomic DNA.
DR Biocyc; RSOL305:RSP0120-MONOMER; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0003994; F:aconitate hydratase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR00573; Aconitase C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF0694; Aconitase; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhyd; 1.
KW Complete proteome; Lyase; plasmid; Transmembrane.
SQ SEQUENCE 865 AA; 93731 MW; 9D59EB071A2D2004 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 865;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
Db 406 CTNTSNPR 413

RESULT 32
Q8XT18_RALSO
ID Q8XT18_RALSO PRELIMINARY; PRT; 865 AA.
AC Q8XT18;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2002, sequence version 1.
DT 07-FEB-2006, entry version 16.
DE PROBABLE ACONITATE HYDRATASE 1 TRANSMEMBRANE PROTEIN (EC 4.2.1.3).
GN Name=acnA2; OrderedLocusNames=RSP0120; ORFNames=RS03002;
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1];
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=GM11009;
RX MEDLINE=21681879; PubMed=11823852; DOI=10.1038/415497a;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Catolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siguiet P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
EMBL; AL646076; CAD17271.1; -; Genomic DNA.
DR Biocyc; RSOL305:RSP0120-MONOMER; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0003994; F:aconitate hydratase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR00573; Aconitase C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF0694; Aconitase; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhyd; 1.
KW Complete proteome; Lyase; plasmid; Transmembrane.
SQ SEQUENCE 865 AA; 93731 MW; 9D59EB071A2D2004 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 865;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
Db 406 CTNTSNPR 413

```

```

RT 3."
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ANA-3;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome of Shewanella sp. ANA-3.";
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AALH01000012; EAP19913.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
SQ SEQUENCE 867 AA; 94752 MW; 06C78A219E1371D7 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 867;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 8
Db 410 CTNTSNPR 417

RESULT 35
Q3P203 9GAMM
ID Q3P203_9GAMM PRELIMINARY; PRT; 867 AA.
AC Q3P203;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, C-terminal; Aconitate hydratase, N-terminal.
GN ORFNames=SDendRAFT_0483;
OS Shewanella denitrificans OS217.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=318161;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OS-217;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Shewanella
RT denitrificans OS-217.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OS-217;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Shewanella denitrificans
RT OS-217.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AALU01000010; EAN70542.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
```

```

RT 3."
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ANA-3;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome of Shewanella sp. ANA-3.";
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AALH01000012; EAP19913.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
SQ SEQUENCE 867 AA; 94742 MW; 06C78A219E1371D7 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 867;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTNTDNP 8
Db 410 CTNTSNPR 417

RESULT 35
Q3P203 9GAMM
ID Q3P203_9GAMM PRELIMINARY; PRT; 867 AA.
AC Q3P203;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, C-terminal; Aconitate hydratase, N-terminal.
GN ORFNames=SDendRAFT_0483;
OS Shewanella denitrificans OS217.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=318161;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OS-217;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Shewanella
RT denitrificans OS-217.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OS-217;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Shewanella denitrificans
RT OS-217.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AALU01000010; EAN70542.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
```

```
DR GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR000573; Aconitase C.
DR InterPro; IPR001030; Aconitase N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
SQ SEQUENCE 867 AA; 94825 MW; 9867049521DCB0A0 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 867;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
Db 406 CTNTSNPR 413

RESULT 36
Q4J5X3 AZOVI
ID Q4J5X3 AZOVI PRELIMINARY; PRT; 867 AA.
AC Q4J5X3;
DT 16-AUG-2005, integrated into UniProtKB/TREMBL.
DT 16-AUG-2005, sequence version 1.
DT 07-FEB-2006, entry version 2.
DE Aconitate hydratase, C-terminal; Aconitate hydratase, N-terminal.
DE ORFNames=AVINDRAFT_5705;
GN Azotobacter vinelandii AVOP.
OS Azotobacter vinelandii AVOP.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=322710;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AVOP;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome assembly of Azotobacter vinelandii
RT AVOP."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AVOP;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Azotobacter vinelandii
RT AVOP."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AVOP;
RG US DOE Joint Genome Institute;
RA DOE Joint Genome Institute;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AVOP;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AAU03000001; EAM07018.1; -; Genomic DNA.
DR GO:0016836; F:hydro-lyase activity; IEA.
DR GO:0005506; F:iron ion binding; IEA.
DR GO:0016829; P:hydro-lyase activity; IEA.
DR GO:0005506; F:iron ion binding; IEA.
DR GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR001030; Aconitase N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
SQ SEQUENCE 867 AA; 94721 MW; D33BB751213AB8F4 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 867;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
Db 413 CTNTSNPR 420

RESULT 37
Q8EJW3 SHEON
ID Q8EJW3 SHEON PRELIMINARY; PRT; 867 AA.
AC Q8EJW3;
DT 01-MAR-2003, integrated into UniProtKB/TREMBL.
DT 01-MAR-2003, sequence version 1.
DT 21-FEB-2006, entry version 13.
DE Aconitate hydratase 1.
DE ORFNames=SO_0343;
GN Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=MR-1;
RG MEDLINE=22297686; PubMed=12368813; DOI=10.1038/nbt749;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N.L., Methe B.A.,
RA Clayton R.A., Meyer I., Tsapin A., Scott J., Beanan M.J.,
RA Brinkac L.M., Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S.,
RA Haft D.H., Kolonay J.F., Madupu R., Peterson J.D., Umayam L.A.,
RA White O., Wolf A.M., Vamathevan J.J., Weidman J.F., Imbraim M.,
RA Lee K., Berry K.J., Lee C., Mueller J., Khouri H.M., Gill J.,
RA Utterback T.R., McDonald L.A., Feldblyum T.V., Smith H.O.,
RA Venter J.C., Nealon K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AE014299; AAN53428.1; -; Genomic DNA.
DR BioCyc; SONE211586:SO0343-MONOMER; -.
DR GO:0016836; F:hydro-lyase activity; IEA.
DR GO:0005506; F:iron ion binding; IEA.
DR GO:0005506; F:iron ion binding; IEA.
DR GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dhyd.
DR InterPro; IPR001030; Aconitase N.
DR InterPro; IPR000573; Aconitase C.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dhy; 1.
Complete proteome.
KW
```

```

SQ SEQUENCE 867 AA; 94596 MW; DA7DBA64A0864A1F CRC64;
Query Match 70.5%; Score 43; DB 2; Length 867;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 410 CTNTSNP 417

RESULT 38
Q3CT85 ALTAT PRELIMINARY; PRT; 868 AA.
AC Q3CT85;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, C-terminal; Aconitate hydratase, N-terminal.
GN ORFNames=PatIDRAFT_3346;
OS Pseudoalteromonas atlantica T6c.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Pseudoalteromonadales; Pseudoalteromonas.
OX NCBI_TaxID=342610;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=T6c;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A.; Lucas S.; Lapidus A.; Barry K.; Detter J.C.; Glavina T.;
RA Hammon N.; Israni S.; Pitluck S.; Richardson P.;
RA "Sequencing of the draft genome and assembly of Pseudoalteromonas
RT atlantica T6c.";
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=T6c;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F.; Land M.;
RA "Annotation of the draft genome assembly of Pseudoalteromonas
RT atlantica T6c.";
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
-!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAKP0100011; EAO68319.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
SQ SEQUENCE 868 AA; 95594 MW; 764CAC26C31D9B5E CRC64;

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 406 CTNTSNP 413

RESULT 39
Q48017 COLP3 PRELIMINARY; PRT; 868 AA.
AC Q48017;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Putative 2-methyl-cis-aconitic acid hydratase.
GN OrderedLocusNames=CPS_2820;

```

```

OS Colwellia psychrerythraea (strain 34H / ATCC BAA-681) (Vibrio
OS psychrerythrus).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Colwelliaceae; Colwellia.
OX NCBI_TaxID=167879;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=16043709; DOI=10.1073/pnas.0504766102;
RA Methe B.A.; Nelson K.E.; Deming J.W.; Momen B.; Melamud E.; Zhang X.;
RA Moulton J.; Madupu R.; Nelson W.C.; Dodson R.J.; Brinkac L.M.;
RA Daugherty S.C.; Durkin A.S.; DeBoy R.T.; Kolonay J.F.; Sullivan S.A.;
RA Zhou L.; Davidson T.M.; Wu M.; Huston A.L.; Lewis M.; Weaver B.;
RA Weidman J.F.; Khouri H.; Utterback T.R.; Feldblyum T.V.; Fraser C.M.;
RT "The psychrophilic lifestyle as revealed by the genome sequence of
RT Colwellia psychrerythraea 34H through genomic and proteomic
RT analyses.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:10913-10918 (2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP0000083; AAZ25694.1; -; Genomic DNA.
DR TIGR; CPS_2820; -;
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met_isocit_dHYD.
DR InterPro; IPR012084; Aco_LysF.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LysF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRFAMs; TIGR02333; 2met_isocit_dHY; 1.
KW Complete proteome.
SQ SEQUENCE 868 AA; 94999 MW; FDC0F3111617B0E CRC64;

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 411 CTNTSNP 418

RESULT 40
Q7NWD6 CHRVO PRELIMINARY; PRT; 868 AA.
AC Q7NWD6;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 13.
DE Aconitate hydratase (EC 4.2.1.3).
GN Name=acn2; OrderedLocusNames=CV2054; ORFNames=CV_2054;
OS Chromobacterium violaceum.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Chromobacterium.
OX NCBI_TaxID=5336;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 12472 / DSM 30191;
RX MEDLINE=22882880; PubMed=14500782; DOI=10.1073/pnas.1832124100;
RA Vasconcelos A.T.R.; de Almeida D.F.; Hungria M.; Guimaraes C.T.;
RA Antonio R.V.; Almeida F.C.; de Almeida L.G.P.; de Almeida R.;
RA Alves-Gomes J.A.; Andrade E.M.; Araripe J.; de Araujo M.F.F.;
RA Astolfi-Filho S.; Azevedo V.; Baptista A.J.; Batais L.A.M.;
RA Batista J.S.; Belo A.; van den Berg C.; Bogo M.; Bonatto S.;
RA Bordignon J.; Brigido M.M.; Brito C.A.; Brocchi M.; Burity H.A.;
RA Camargo A.A.; Cardoso D.P.; Carneiro N.P.; Carraro D.M.;
RA Carvalho C.M.B.; Cascardo J.C.M.; Cavada B.S.; Chueire L.M.O.;

```

RA Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,  
RA Fantinatti F., Parias I.P., Felipe M.S.A., Ferrari L.P., Ferro J.A.,  
RA Ferro M.I.T., Franco G.R., Freitas N.S.A., Furlan L.R.,  
RA Gazzinelli R.T., Gomes E.A., Goncalves P.R., Grangeiro T.B.,  
RA Grattapaglia D., Grissard E.C., Hanna E.S., Jardim S.N., Laurino J.,  
RA Leoi L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,  
RA Madeira H.M.F., Manfio G.P., Maranhao A.Q., Martins W.S.,  
RA di Mauro S.M.Z., de Medeiros S.R.B., Meissner R.V., Moreira M.A.M.,  
RA Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,  
RA Paixao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,  
RA Ramalho M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,  
RA Remeiro-Neto C.B., Reis A.M.M., Rigo L.U., Rondinelli E.,  
RA Santos E.B.P., Santos F.R., Schneider M.P.C., Seanez H.N.,  
RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,  
RA Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,  
RA Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Urmenyi T.,  
RA Vettore A., Wassem R., Zaha A., Simpson A.J.G.;  
RT "The complete genome sequence of Chromobacterium violaceum reveals  
RT remarkable and exploitable bacterial adaptability.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; AE016825; AA059726.1; -; Genomic\_DNA.  
DR BioCyc; CVI0243365; CV2054-MONOMER; -;  
DR GO; GO:0003994; F:aconitate hydratase activity; IEA.  
DR GO; GO:0005506; F:iron ion binding; IEA.  
DR GO; GO:0016829; F:lyase activity; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR012708; 2met\_isocit\_dHVD.  
DR InterPro; IPR012084; AcoLysF.  
DR InterPro; IPR000573; Aconitase C.  
DR InterPro; IPR001030; Aconitase\_N.  
DR Pfam; PF00330; Aconitase; 1.  
DR Pfam; PF00694; Aconitase\_C; 1.  
DR PIRSF; PIRSF001417; LysF; 1.  
DR PRINTS; PR00415; ACONITASE.  
DR ProDom; PD000511; Aconitase\_N; 1.  
DR TIGRFAMs; TIGR02333; 2met\_isocit\_dHY; 1.  
KW Complete proteome; Lyase.  
SQ SEQUENCE 868 AA; 94648 MW; D0FB836453F66F95 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 868;  
Best Local Similarity 87.5%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8  
|||||  
DB 411 CTNTSNPR 418

RESULT 41  
Q915E4 PSBAE PRELIMINARY; PRT; 868 AA.  
AC Q915E4;  
DT 01-MAR-2001, integrated into UniProtKB/TrEMBL.  
DT 01-FEB-2006, entry version 16.  
DE Probable aconitate hydratase.  
GN OrderedLocusNames=PA0794;  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=ATCC 15692 / PA01.  
RX MEDLINE=20437337; PubMed=10984043; DOI=10.1038/35023079;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H. Jr., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; AE004514; AAG04183.1; -; Genomic\_DNA.  
DR PIR; AB3547; AB3547.  
DR BioCyc; PAER287; PA0794-MONOMER; -;  
DR LinkHub; Q915E4; -;  
DR GO; GO:0016836; F:hydro-lyase activity; IEA.  
DR GO; GO:0005506; F:iron ion binding; IEA.  
DR GO; GO:0016829; F:lyase activity; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR012708; 2met\_isocit\_dHVD.  
DR InterPro; IPR012084; AcoLysF.  
DR InterPro; IPR000573; Aconitase C.  
DR InterPro; IPR001030; Aconitase\_N.  
DR Pfam; PF00330; Aconitase; 1.  
DR Pfam; PF00694; Aconitase\_C; 1.  
DR PIRSF; PIRSF001417; LysF; 1.  
DR PRINTS; PR00415; ACONITASE.  
DR ProDom; PD000511; Aconitase\_N; 1.  
DR TIGRFAMs; TIGR02333; 2met\_isocit\_dHY; 1.  
KW Complete proteome.  
SQ SEQUENCE 868 AA; 94891 MW; 5A6E0C9D30EC328F CRC64;

Query Match 70.5%; Score 43; DB 2; Length 868;  
Best Local Similarity 87.5%; Pred. No. 82;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8  
|||||  
DB 411 CTNTSNPR 418

RESULT 42  
Q9JT05 NEIMA PRELIMINARY; PRT; 868 AA.  
AC Q9JT05;  
DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.  
DT 01-OCT-2000, sequence version 1.  
DT 07-FEB-2006, entry version 19.  
DE Aconitate hydratase (EC 4.2.1.3).  
GN Name=acnA; OrderedLocusNames=NMA2052;  
OS Neisseria meningitidis serogroup A.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
OC Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=65699;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=Z2491 / Serogroup A / Serotype 4A;  
RX MEDLINE=20222556; PubMed=10761919; DOI=10.1038/35006655;  
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.M.,  
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,  
RA Jagels K., Leather S., Moule S., Mungall K.L., Quail M.A.,  
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
RA Whitehead S., Spratt B.G., Barrall B.G.;  
RT "Complete DNA sequence of a serogroup A strain of Neisseria  
RT meningitidis Z2491";  
RL Nature 404:502-506(2000).  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; AL162758; CAB85270.1; -; Genomic\_DNA.  
DR PIR; H81775; H81775.  
DR BioCyc; NMN65699; NMA2052-MONOMER; -;  
DR GO; GO:0003994; F:aconitate hydratase activity; IEA.



```
DR GO: GO:0005506; F:iron ion binding; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR012708; 2met_isocit_dhyd.
DR InterPro: IPR012084; Aco_LysF.
DR InterPro: IPR000573; Aconitase_C.
DR InterPro: IPR001030; Aconitase_N.
DR Pfam: PF00330; Aconitase; 1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PIRSF: PIRSF001417; LysF; 1.
DR PRINTS: PR00415; ACONITASE.
DR ProDom: PD000511; Aconitase_N; 1.
DR TIGRFAMs: TIGR02333; 2met_isocit_dhy; 1.
KW Complete proteome; Lyase.
SQ SEQUENCE 868 AA; 95030 MW; 0D7FE359E772F9A8 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 412 CTNTSNPR 419

RESULT 43
Q9K0X3 NEIMB PRELIMINARY; PRT; 868 AA.
AC Q9K0X3;
DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2000, sequence version 1.
DT 07-FEB-2006, entry version 20.
DE Aconitate hydratase 1 (EC 4.2.1.3).
GN Name: acoA; OrderedLocNames: NM004033;
OS Neisseria meningitidis serogroup B.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=MC58 / Serogroup B.
RX MEDLINE=20175755; PubMed=10710307; DOI=10.1126/science.287.5459.1809;
RA Tettelin H., Saunders N.J., Heidelberg J.F., Jeffries A.C.,
RA Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F.,
RA Dodson R.J., Nelson M.C., Gwinn M.L., DeBoy R.T., Peterson J.D.,
RA Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D.,
RA Dougherty B.A., Mason T.M., Ciecho A., Parksey D.S., Blair E.,
RA Cittone H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H.M.,
RA Qin H., Vamathevan J.J., Gill J., Scarlato V., Maignani V., Pizza M.,
RA Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R.,
RA Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815 (2000).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
EMBL: AB020098; AAF40871.1; -, Genomic_DNA.
PIR: C81200; C81200.
TIGR: NMB0433; -.
BioCyc: NMEN491:NMB0433-MONOMER; -.
DR GO: GO:0003994; F:aconitase hydratase activity; IEA.
DR GO: GO:0005506; F:iron ion binding; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR012708; 2met_isocit_dhyd.
DR InterPro: IPR012084; Aco_LysF.
DR InterPro: IPR000573; Aconitase_C.
DR InterPro: IPR001030; Aconitase_N.
DR Pfam: PF00330; Aconitase; 1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PIRSF: PIRSF001417; LysF; 1.

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 412 CTNTSNPR 419

RESULT 44
Q9KSC0 VIBCH PRELIMINARY; PRT; 868 AA.
AC Q9KSC0;
DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2000, sequence version 1.
DT 07-FEB-2006, entry version 17.
DE Aconitate hydratase 1.
GN OrderedLocNames: VCL338;
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301; DOI=10.1038/35020000;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.L.,
RA Ermolaeva M.D., Vamathevan J.J., Bass S., Qin H., Dragoi I.,
RA Sellers P., McDonald L.A., Utterback T.R., Fleischmann R.D.,
RA Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R.,
RA Mekalanos J.J., Venter J.C., Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483 (2000).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
EMBL: AE004213; AAF94496.1; -, Genomic_DNA.
PIR: B82213; B82213.
TIGR: VCI338; -.
DR GO: GO:0016836; F:hydro-lyase activity; IEA.
DR GO: GO:0005506; F:iron ion binding; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR012708; 2met_isocit_dhyd.
DR InterPro: IPR012084; Aco_LysF.
DR InterPro: IPR000573; Aconitase_C.
DR InterPro: IPR001030; Aconitase_N.
DR Pfam: PF00330; Aconitase; 1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PIRSF: PIRSF001417; LysF; 1.
DR PRINTS: PR00415; ACONITASE.
DR ProDom: PD000511; Aconitase_N; 1.
DR TIGRFAMs: TIGR02333; 2met_isocit_dhy; 1.
KW Complete proteome.
SQ SEQUENCE 868 AA; 95073 MW; E8837B65E9FFED CRC64;

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 415 CTNTSNPR 422
```

```
DR PRINTS: PR00415; ACONITASE.
DR ProDom: PD000511; Aconitase_N; 1.
DR TIGRFAMs: TIGR02333; 2met_isocit_dhy; 1.
KW Complete proteome; Lyase.
SQ SEQUENCE 868 AA; 95097 MW; 1179C795F538B037 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 412 CTNTSNPR 419

RESULT 44
Q9KSC0 VIBCH PRELIMINARY; PRT; 868 AA.
AC Q9KSC0;
DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2000, sequence version 1.
DT 07-FEB-2006, entry version 17.
DE Aconitate hydratase 1.
GN OrderedLocNames: VCL338;
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301; DOI=10.1038/35020000;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.L.,
RA Ermolaeva M.D., Vamathevan J.J., Bass S., Qin H., Dragoi I.,
RA Sellers P., McDonald L.A., Utterback T.R., Fleischmann R.D.,
RA Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R.,
RA Mekalanos J.J., Venter J.C., Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483 (2000).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
EMBL: AE004213; AAF94496.1; -, Genomic_DNA.
PIR: B82213; B82213.
TIGR: VCI338; -.
DR GO: GO:0016836; F:hydro-lyase activity; IEA.
DR GO: GO:0005506; F:iron ion binding; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR012708; 2met_isocit_dhyd.
DR InterPro: IPR012084; Aco_LysF.
DR InterPro: IPR000573; Aconitase_C.
DR InterPro: IPR001030; Aconitase_N.
DR Pfam: PF00330; Aconitase; 1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PIRSF: PIRSF001417; LysF; 1.
DR PRINTS: PR00415; ACONITASE.
DR ProDom: PD000511; Aconitase_N; 1.
DR TIGRFAMs: TIGR02333; 2met_isocit_dhy; 1.
KW Complete proteome.
SQ SEQUENCE 868 AA; 95073 MW; E8837B65E9FFED CRC64;

Query Match 70.5%; Score 43; DB 2; Length 868;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
Db 415 CTNTSNPR 422
```

```
RESULT 45
Q31H91_PSEHT
ID Q31H91_PSEHT PRELIMINARY; PRT; 869 AA.
AC Q31H91;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 7.
DE Aconitate hydratase 1 (EC 4.2.1.3).
GN Name=acnA; OrderedLocNames=PSHAa1773; ORFNames=PSHAa1773;
OS Pseudocalteromonas haloplanktis (strain TAC 125).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Pseudocalteromonadaceae; Pseudocalteromonas.
OX NCBI_TaxID=326442;
RN [1]
NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=16169927; DOI=10.1101/gr.4126905;
RA Medigue C., Krin E., Pascal G., Barbe V., Bernsel A., Bertin P.N.,
RA Cheung F., Cruveiller S., D'Amico S., Duilio A., Fang G., Feller G.,
RA Ho C., Mangerot S., Marino G., Nilsson J., Parrilli E., Rocha E.P.C.,
RA Rouy Z., Sekowska A., Tutino M.L., Vallenet D., von Heijne G.,
RA Danchin A.;
RT "Coping with cold: the genome of the versatile marine Antarctic
RT bacterium Pseudocalteromonas haloplanktis TAC125.";
RL Genome Res. 15:1325-1335(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL: CR954246; CAI86845.1; -: Genomic DNA.
DR GO: GO:0003994; F:aconitate hydratase activity; IEA.
DR GO: GO:0005062; F:iron ion binding; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR012708; 2met_isocit_dhyd.
DR InterPro: IPR012084; Aco_LysF.
DR InterPro: IPR000573; Aconitase_C.
DR InterPro: IPR001030; Aconitase_N.
DR Pfam: PF00330; Aconitase; 1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PIRSF: PIRSF001417; LysF; 1.
DR PRINTS: PR00415; ACONITASE.
DR ProDom: PD000511; Aconitase_N; 1.
DR TIGRFAMs: TIGR02333; 2met_isocit_dhy; 1.
KW Complete proteome; Lyase.
SQ SEQUENCE 869 AA; 95141 MW; EB6A02459E658899 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 869;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
Db 410 CTNTSNPR 417

RESULT 46
Q937N8_RALEU
ID Q937N8_RALEU PRELIMINARY; PRT; 869 AA.
AC Q937N8;
DT 01-DEC-2001, integrated into UniProtKB/TrEMBL.
DT 01-DEC-2001, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Probable methyl-cis-aconitic acid hydratase.
GN Name=acnM;
OS Ralstonia eutropha (Alcaligenes eutrophus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Cupriavidus.
OX NCBI_TaxID=106590;
RN [1]
NUCLEOTIDE SEQUENCE.
RC STRAIN=HF39;
```

```
RX MEDLINE=21387530; PubMed=11495997;
RA Bramer C.O., Steinbuechel A.;
RT "The methylicitric acid pathway in Ralstonia eutropha: new genes
RT identified involved in propionate metabolism.";
RL Microbiology 147:2203-2214(2001).
RN [2]
NUCLEOTIDE SEQUENCE.
RC STRAIN=HF39;
RA Bramer C.O., Steinbuechel A.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL: AF325554; AAL03990.1; -: Genomic DNA.
DR GO: GO:0016836; F:hydro-lyase activity; IEA.
DR GO: GO:0005062; F:iron ion binding; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR012708; 2met_isocit_dhyd.
DR InterPro: IPR012084; Aco_LysF.
DR InterPro: IPR000573; Aconitase_C.
DR InterPro: IPR001030; Aconitase_N.
DR Pfam: PF00330; Aconitase; 1.
DR Pfam: PF00694; Aconitase_C; 1.
DR PIRSF: PIRSF001417; LysF; 1.
DR PRINTS: PR00415; ACONITASE.
DR ProDom: PD000511; Aconitase_N; 1.
DR TIGRFAMs: TIGR02333; 2met_isocit_dhy; 1.
SQ SEQUENCE 869 AA; 94726 MW; DA11CA78DD0C9710 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 869;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
Db 411 CTNTSNPR 418

RESULT 47
Q2KW75_BORAV
ID Q2KW75_BORAV PRELIMINARY; PRT; 869 AA.
AC Q2KW75;
DT 07-MAR-2006, integrated into UniProtKB/TrEMBL.
DT 07-MAR-2006, sequence version 1.
DT 07-MAR-2006, entry version 1.
DE Aconitate hydratase 2 (EC 4.2.1.3).
GN Name=acnA4; ORFNames=BAV2732;
OS Bordetella avium 197N.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OX Alcaligenaceae; Bordetella.
RN [1]_TaxID=360910;
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=197N;
RA Sebaihia M.;
RT "The genome sequence of the poultry pathogen Bordetella avium, and
RT genomic comparisons with related species infecting mammals.";
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL: AM167904; CAJ50343.1; -: Genomic DNA.
DR Lyase.
SQ SEQUENCE 869 AA; 95233 MW; 0BC85A015BE50385 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 869;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNR 8
```

```
Db 406 CTNTSNPR 413
|||||
RESULT 48
Q3NW39_SHEFR PRELIMINARY; PRT; 870 AA.
AC Q3NW39;
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, N-terminal.
GN ORFNames=Sfr1DRAFT_3713;
OS Shewanella frigidimarina NCIMB 400.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=318167;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCIMB 400;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Iseri S., Pittluck S., Richardson P.,
RT "Sequencing of the draft genome and assembly of Shewanella
frigidimarina NCIMB 400."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCIMB 400;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Shewanella frigidimarina
NCIMB 400."
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AA101000001; EAF75929.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met isocit dHYD.
DR InterPro; IPR012084; AcoLySF.
DR InterPro; IPR000573; Aconitase C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LySF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRams; TIGR02333; 2met isocit dHY; 1.
SQ SEQUENCE 870 AA; 95629 MW; 612651CBF03D41EF CRC64;

Query Match 70.5%; Score 43; DB 2; Length 870;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
|||||
Db 413 CTNTSNPR 420

RESULT 49
Q2WXY8_9GAMM PRELIMINARY; PRT; 871 AA.
AC Q2WXY8;
DT 10-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 10-JAN-2006, entry version 1.
DE Aconitate hydratase, N-terminal.
GN ORFNames=Sfr1DRAFT_3713;
OS Shewanella frigidimarina NCIMB 400.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=318167;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCIMB 400;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Iseri S., Pittluck S., Richardson P.,
RT "Sequencing of the draft genome and assembly of Shewanella
frigidimarina NCIMB 400."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCIMB 400;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Shewanella frigidimarina
NCIMB 400."
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AA101000001; EAF75929.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met isocit dHYD.
DR InterPro; IPR012084; AcoLySF.
DR InterPro; IPR000573; Aconitase C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LySF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRams; TIGR02333; 2met isocit dHY; 1.
SQ SEQUENCE 870 AA; 95629 MW; 612651CBF03D41EF CRC64;

Query Match 70.5%; Score 43; DB 2; Length 870;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
|||||
Db 413 CTNTSNPR 420

RESULT 49
Q2WXY8_9GAMM PRELIMINARY; PRT; 871 AA.
AC Q2WXY8;
DT 10-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 10-JAN-2006, entry version 1.
DE Aconitate hydratase, N-terminal.
GN ORFNames=Sfr1DRAFT_3713;
OS Shewanella frigidimarina NCIMB 400.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=318167;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCIMB 400;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hammon N., Iseri S., Pittluck S., Richardson P.,
RT "Sequencing of the draft genome and assembly of Shewanella
frigidimarina NCIMB 400."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCIMB 400;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Shewanella frigidimarina
NCIMB 400."
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AA101000001; EAF75929.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met isocit dHYD.
DR InterPro; IPR012084; AcoLySF.
DR InterPro; IPR000573; Aconitase C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LySF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRams; TIGR02333; 2met isocit dHY; 1.
SQ SEQUENCE 870 AA; 95629 MW; 612651CBF03D41EF CRC64;
```

```
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, N-terminal.
GN ORFNames=Sputw3181DRAFT_1044;
OS Shewanella sp. W3-18-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=351745;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W3-18-1;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Iseri S., Pittluck S., Richardson P.,
RT "Sequencing of the draft genome and assembly of Shewanella sp. W3-18-
1."
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W3-18-1;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Shewanella sp. W3-18-1."
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AALN01000121; EAP53767.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met isocit dHYD.
DR InterPro; IPR012084; AcoLySF.
DR InterPro; IPR000573; Aconitase C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LySF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRams; TIGR02333; 2met isocit dHY; 1.
SQ SEQUENCE 871 AA; 95177 MW; B815E0DB1FB039E7 CRC64;

Query Match 70.5%; Score 43; DB 2; Length 871;
Best Local Similarity 87.5%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTNTDNP 8
|||||
Db 414 CTNTSNPR 421

RESULT 50
Q2ZVF7_SHEPU PRELIMINARY; PRT; 871 AA.
AC Q2ZVF7;
DT 20-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 20-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Aconitate hydratase, C-terminal; Aconitase hydratase, N-terminal.
GN ORFNames=Sputcn32DRAFT_2935;
OS Shewanella putrefaciens CN-32.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=319224;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CN-32;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Iseri S., Pittluck S., Richardson P.,
RT "Sequencing of the draft genome and assembly of Shewanella
putrefaciens CN-32."
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CN-32;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome of Shewanella putrefaciens CN-32."
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AALN01000121; EAP53767.1; -; Genomic DNA.
DR GO; GO:0016836; F:hydro-lyase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR012708; 2met isocit dHYD.
DR InterPro; IPR012084; AcoLySF.
DR InterPro; IPR000573; Aconitase C.
DR InterPro; IPR001030; Aconitase_N.
DR Pfam; PF00330; Aconitase; 1.
DR Pfam; PF00694; Aconitase_C; 1.
DR PIRSF; PIRSF001417; LySF; 1.
DR PRINTS; PR00415; ACONITASE.
DR ProDom; PD000511; Aconitase_N; 1.
DR TIGRams; TIGR02333; 2met isocit dHY; 1.
SQ SEQUENCE 870 AA; 95629 MW; 612651CBF03D41EF CRC64;
```

```

CC      preliminary data.
CC      -----
CC      Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC      Distributed under the Creative Commons Attribution-NoDerivs License
CC      -----
DR      EMBL; AALB01000004; EAO95942.1; -; Genomic DNA.
SQ      SEQUENCE 871 AA; 95191 MW; 470199BA8368C1F3 CRC64;

Query Match      70.5%; Score 43; DB 2; Length 871;
Best Local Similarity 87.5%; Pred. NO. 82;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CTNTDNPR 8
      |||||
Db      414 CTNTSNPR 421

Search completed: June 5, 2006, 12:54:32
Job time : 121.671 secs

```

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:06 ; Search time 104.268 Seconds  
(without alignments)  
2293.354 Million cell updates/sec

Title: US-10-645-659A-5  
Perfect score: 2728  
Sequence: 1 MEVLILLVLLAVPPRRTRAE.....LPAFSYGFYVRNAKAIACI 523

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_8:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*  
9: Geneseqp2005s:\*  
10: Geneseqp2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2728	100.0	523	5 ABB07814	Chicken h
2	2728	100.0	523	7 ABW02017	Chicken h
3	2728	100.0	523	8 ADR88211	Chicken h
4	2728	100.0	523	8 ADT78178	Chicken h
5	2728	100.0	523	9 ADY27037	Chicken h
6	2728	100.0	523	9 AEA42427	Chicken h
7	1717	62.9	527	5 ABB07815	Chicken s
8	1717	62.9	527	7 ABW02018	Chimeric
9	1711	62.7	527	8 ADO63825	Chimeric
10	1711	62.7	527	8 ADO63826	Chimeric
11	1705	62.5	527	8 ADO63827	Chimeric
12	1653.5	60.6	543	8 ADO63832	Human hep
13	1650	60.5	545	6 ABP56822	Human hep
14	1650	60.5	545	7 ADE16012	G-coupled
15	1650	60.5	545	8 ADL93951	Human G-c
16	1648.5	60.4	543	2 AAY17082	Human hep
17	1648.5	60.4	543	4 AAB86206	Human hep
18	1648.5	60.4	543	7 ADD18950	Human dis
19	1648.5	60.4	543	8 ADK52086	Human ato
20	1648.5	60.4	543	8 ADM48759	Human hpa
21	1648.5	60.4	543	8 ADN05074	Antipori
22	1648.5	60.4	543	8 ADN04902	Antipori
23	1648.5	60.4	543	8 ADQ80372	Heparanas

24	1648.5	60.4	543	8 ADR88210	Human pre
25	1648.5	60.4	543	8 ADP25079	PRO polyep
26	1648.5	60.4	543	8 ADT78177	Human hep
27	1648.5	60.4	543	9 ADY27036	Human hep
28	1648.5	60.4	543	9 AEA42426	Human hep
29	1648.5	60.4	588	2 AAY30124	A human h
30	1648.5	60.3	543	2 AAY02345	A human h
31	1645.5	60.3	543	3 AAY57590	Human hep
32	1645.5	60.3	543	3 AAB08849	Amino aci
33	1645.5	60.3	543	3 AAY52990	Human hep
34	1645.5	60.3	543	4 AAY97635	Human hep
35	1645.5	60.3	543	5 ABB07813	Human hep
36	1645.5	60.3	543	7 ADG88800	Human hpa
37	1645.5	60.3	543	8 ADL16379	Human hep
38	1645.5	60.3	543	8 ADM48716	Human hpa
39	1645.5	60.3	543	9 AEA42466	Human hep
40	1645.5	60.3	543	10 AEE96848	Human hep
41	1645.5	60.3	592	2 AAY02346	A human h
42	1645.5	60.3	592	3 AAB08850	Amino aci
43	1645.5	60.3	592	7 ADG88804	Human SK-
44	1645.5	60.3	592	8 ADL16383	Human hep
45	1645.5	60.3	592	8 ADM48720	Human SK-

ALIGNMENTS

RESULT 1  
ABB07814  
ID ABB07814 standard; protein; 523 AA.  
XX  
AC ABB07814;  
XX  
DT 03-JUL-2002 (first entry)  
XX  
DE Chicken heparanase sequence.  
XX  
KW Heparanase; catalytic; cytosolic; antiviral; antibacterial; enzyme;  
KW anti-protozoan; neuroprotective; heparin; chicken.  
XX  
OS Gallus gallus.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..19  
FT Protein /note= "putative signal peptide"  
FT /note= "mature protein"  
XX  
FN US2002034810-A1.  
XX  
PD 21-MAR-2002.  
XX  
PF 16-AUG-2001; 2001US-00930218.  
XX  
PR 20-SEP-2000; 2000US-00666390.  
XX  
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
XX  
PI Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;  
XX  
DR WPI; 2002-338926/37.  
XX  
DR N-PSDB; ABL40748.  
XX  
PT Nucleic acid encoding avian and reptile heparanase polypeptide is useful  
PT to treat various heparin-related disorders and the signal peptide is  
PT useful in production of membrane-targeted or secreted recombinant  
PT proteins.  
XX  
PS Claim 19; Fig 1b; 39pp; English.  
XX  
CC The invention relates to an isolated avian and reptile nucleic acid,  
CC encoding a polypeptide with heparanase catalytic activity. The signal  
CC peptide of the nucleic acid can be used to express membrane-associated or



Db 361 MRQVSFGAGSYHLVDAGFKLPDYWLSLLYKRLVGTTRVLQASVEQADARRPRVYLHCTNP 420  
 QY 421 RHPKYREGDVTLPALNLSNVTQSLQPKQLWSKSVDOYVLLPHGKDSILSREVQLNGRLL 480  
 Db 421 RHPKYREGDVTLPALNLSNVTQSLQPKQLWSKSVDOYVLLPHGKDSILSREVQLNGRLL 480  
 QY 481 QMVDDETLPALHEMALAPGSTLGLPAFSYGYFYVIRNAKAIACI 523  
 Db 481 QMVDDETLPALHEMALAPGSTLGLPAFSYGYFYVIRNAKAIACI 523

## RESULT 3

ADR88211

ID ADR88211 standard; protein; 523 AA.

AC ADR88211;

XX ADR88211;

DT 18-NOV-2004 (first entry)

XX Chicken heparanase.

XX Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;

KW autoimmune disorder; cancer; angiogenesis; metastatic disease;

KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;

KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;

KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;

KW chicken; heparanase; enzyme.

XX Gallus gallus.

OS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

of a patient, the tissue expressing heparanase. The method comprises providing a complex of a drug directly or indirectly linked to an anti-heparanase antibody, and administering the complex to the patient. In the targeted drug delivery, the antibody comprises an epitope of a heparanase capable of specifically binding to at least one epitope of a heparanase protein. The composition and methods of the invention are useful for diagnosing, preventing or treating conditions associated with heparanase catalytic activity (e.g. an inflammatory disorder, wound, scar, vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell proliferation, invasion of circulating tumour cells and metastatic disease), for purifying heparanase, or for developing drugs for those heparanase-associated conditions. The vasculopathy is atherosclerosis, restenosis or aneurysm. The cancerous condition is a solid cancer or a non-solid cancer. The non-solid cancer is a haematopoietic malignancy selected from acute lymphocytic leukaemia (ALL), acute myelogenous leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia, Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and multiple myeloma. The solid cancer is selected from tumours in lip and oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands, thyroid gland, oesophagus, stomach, small intestine, colon, colorectum, anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva, vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva, malignant melanoma of the conjunctiva, malignant melanoma of the uvea, retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit, brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's sarcoma. The present sequence is chicken heparanase.

SQ Sequence 523 AA;

Query Match 100.0%; Score 2728; DB 8; Length 523;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-267;  
 Matches 523; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLVLLLVLLAVPPRTAEQLGLRPIGAVSPAFSLTLTDLASLARDPRFVALLRHPKL 60  
 Db 1 MLVLLLVLLAVPPRTAEQLGLRPIGAVSPAFSLTLTDLASLARDPRFVALLRHPKL 60

QY 61 HTLASGLSPGLRFGTSTDELIENPNKDSWEKVLSEFOAKDYCEAWPSFVVPKLLL 120  
 Db 61 HTLASGLSPGLRFGTSTDELIENPNKDSWEKVLSEFOAKDYCEAWPSFVVPKLLL 120

QY 121 TQWPLQEKLLLAHSHWKKHNTTITRSTLDTLHTFASSSGPRLVFGLNALLRRAGLQWDS 180  
 Db 121 TQWPLQEKLLLAHSHWKKHNTTITRSTLDTLHTFASSSGPRLVFGLNALLRRAGLQWDS 180

QY 181 SNAKQLLAGYCAQRSYNI SWELGNEPNSFRKKSIGICIDGFLGRDFVHLRQLLSQHPLRYH 240  
 Db 181 SNAKQLLAGYCAQRSYNI SWELGNEPNSFRKKSIGICIDGFLGRDFVHLRQLLSQHPLRYH 240

QY 241 AELYGLDVGGPRKHTQHLRSFMKSGGKAIDSVTHHHYVNGRSATREDFISPEVLDSFA 300  
 Db 241 AELYGLDVGGPRKHTQHLRSFMKSGGKAIDSVTHHHYVNGRSATREDFISPEVLDSFA 300

QY 301 TATHDLVIGIVEATVPKKVWLGTSYAGGAPOLSNITYVAGFMWLDKGLAARIGIDVV 360  
 Db 301 TATHDLVIGIVEATVPKKVWLGTSYAGGAPOLSNITYVAGFMWLDKGLAARIGIDVV 360

QY 361 MRQVSFGAGSYHLVDAGFKLPDYWLSLLYKRLVGTTRVLQASVEQADARRPRVYLHCTNP 420  
 Db 361 MRQVSFGAGSYHLVDAGFKLPDYWLSLLYKRLVGTTRVLQASVEQADARRPRVYLHCTNP 420

QY 421 RHPKYREGDVTLPALNLSNVTQSLQPKQLWSKSVDOYVLLPHGKDSILSREVQLNGRLL 480  
 Db 421 RHPKYREGDVTLPALNLSNVTQSLQPKQLWSKSVDOYVLLPHGKDSILSREVQLNGRLL 480

QY 481 QMVDDETLPALHEMALAPGSTLGLPAFSYGYFYVIRNAKAIACI 523  
 Db 481 QMVDDETLPALHEMALAPGSTLGLPAFSYGYFYVIRNAKAIACI 523

Targeted drug delivery to a heparanase-expressing tissue of a patient, useful for treating heparanase-associated conditions such as inflammation or cancer, comprises administering a drug and an anti-heparanase antibody complex.

Claim 2; SEQ ID NO 5; 58pp; English.

The invention relates to a method of targeted drug delivery to a tissue









RESULT 8  
 ID ABW02018  
 ABW02018 standard; protein; 527 AA.  
 AC ABW02018;  
 XX  
 DT 12-FEB-2004 (first entry)  
 XX  
 DE Chimeric human-chicken heparanase protein.  
 XX  
 KW Chicken; heparanase; tumour cell metastasis; inflammation; autoimmunity;  
 KW wound healing; angiogenesis; restenosis; Genstmann-Straussler Syndrome;  
 KW neurodegenerative disease; atherosclerosis; Creutzfeldt-Jakob disease;  
 KW infection; Scrapie; Alzheimer's disease; protein therapy; cytostatic;  
 KW immunosuppressive; vulnery; bactericide; anti-angiogenic; virucide;  
 KW anticlerotic; neuroprotective; protozoacide; chimeric; fusion protein;  
 KW enzyme; human.  
 XX  
 OS Chimeric - Gallus gallus.  
 OS Chimeric - Homo sapiens.  
 XX  
 PN US2003180788-A1.  
 XX  
 PD 25-SEP-2003.  
 XX  
 XX 08-MAY-2003; 2003US-00431438.  
 XX  
 PR 20-SEP-2000; 2000US-00666390.  
 PR 16-AUG-2001; 2001US-00930218.  
 XX  
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
 XX  
 PI Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zoharia E;  
 XX  
 DR WPI; 2003-843931/78.  
 DR N-PSDB; AAD63532.  
 XX  
 PT Recombinant jungle red fowl (Gallus gallus) heparanase protein, useful  
 PT for treating cancers, microbial infections and aiding wound healing.  
 XX  
 PS Example; Page 26-28; Opp; English.  
 XX  
 CC The present invention relates to novel jungle red fowl heparanase protein  
 CC and polynucleotides encoding such proteins. Heparanase sequences can be  
 CC used to develop treatments for various diseases, to develop diagnostic  
 CC assays for these diseases and to provide new tools for basic and directed  
 CC research especially in the fields of medicine and biology. They can be  
 CC used to develop new drugs to inhibit tumour cell metastasis, inflammation  
 CC and autoimmunity. Recombinant heparanase offers a potential treatment for  
 CC wound healing, angiogenesis, restenosis, atherosclerosis, inflammation,  
 CC neurodegenerative diseases (e.g. Genstmann-Straussler Syndrome, Scrapie,  
 CC Creutzfeldt-Jakob disease and Alzheimer's disease) and certain viral and  
 CC some bacterial and protozoa infections. Recombinant heparanase can also  
 CC be used to neutralise plasma heparin, as a potential replacement of  
 CC protamine. Sequences of the invention are used in protein therapy. The  
 CC present sequence is chimeric human-chicken heparanase protein  
 XX  
 SQ Sequence 527 AA;  
 Query Match 62.9%; Score 1717; DB 7; Length 527;  
 Best Local Similarity 63.1%; Pred. No. 9.1e-165;  
 Matches 333; Conservative 83; Mismatches 106; Indels 6; Gaps 3;  
 QY 1 MLVLLLVLLVPPRRRTA---ELQLGREPIGAVSPAFSLTLDASLARDPRFVALLR 56  
 DB 1 MLVLLLVLLVPPRRRTAQDVLDFFTEPLHLVSPFLSVTIDANLATDPRFLILG 60  
 QY 57 HPKLHTLASGLSPGFLRGFTSTDFLIFNPNKOSTWEKVLSEFQA-KDVCAPWSPFAV 115  
 DB 61 SPKRLTLARGLSPAYLRFGGTKTDFLIFDPKKESTFEERSYQSQVNQDICKYGSIPDPV 120

QY 116 PKLLLTQWPLQEKLLLAHSHWKKHNTTITRSTLDILHTFASSGFRLVFGNALLRRAG 175  
 DB 121 EEKLRLEWPEYQEQLLLEHYOKFKPNSTYSRSSVDVLYTFANCSDLDFGNALLRTAD 180  
 QY 176 LQWSSNAKQLLYGCAORSYNISWELGNEPNSFRKSGICIDGFGQFQGRDFVHLRQLLSQH 235  
 DB 181 LQWSSNAQLLLDYCSSKGYNISWELGNEPNSFLKADIFINGSQLGEDFIQLHKL-RK 239  
 QY 236 PLYRHAEYGLDVGOPRKHTOHLRLSRFMKSGGKAIDSVTWHYYVNGRSATREDFLSPV 295  
 DB 240 STFKNALYGPDVQPPRRKTAKMLKSLFKAGGEVIDSVTHYYLNGRTATREDFLNPV 299  
 QY 296 LDSFATAIHDLVGLVEATVPGKVMWLGFTSGAYGGAPQLSNTYYVAGFMWLDKGLAARR 355  
 DB 300 LDIFISSQVKVQFVVESTRPGKVMWLGFTSGAYGGAPQLSNTFAAGFMWLDKGLSARM 359  
 QY 356 GIDVVMQVSGFAGSYHLVDAGFKPLPDYWLISLYKRLVGTGVLOASVEQADARRPVYL 415  
 DB 360 GIEVVMQVFGAGNYHLVDENFDPLPDYWLISLFFKLVGTVKVMASVQSGSKRKLRYL 419  
 QY 416 HCTNPRHPKYREGDVTLPALNLSNVTOSLQLPKQLMWSKSDVOYLLPHGKDSILSREVOL 475  
 DB 420 HCTNTDNPRYKEGDLTLVAINLHNVTYKYLRLPYPSNKQVDKYLLRPLGPHGLLSKVQL 479  
 QY 476 NGRLLQWVDDDELTPALHEMALAPGSTIGLPAPSYGFYVIRNAKAIACI 523  
 DB 480 NGLTLKWVDDQTLPLPMEKPLRPGSSGLPAPSYGFYVIRNAKVAACI 527  
 RESULT 9  
 ADO63825  
 ID ADO63825 standard; protein; 527 AA.  
 XX  
 AC ADO63825;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Chimeric heparanase mutant E225A, SEQ ID:10.  
 XX  
 KW Human; chicken; heparanase; heparanase-derived protein;  
 KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;  
 KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;  
 KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;  
 KW adhesion disorder; IAD; leukocyte adhesion deficiency;  
 KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
 KW vulnery; mutant; mutein.  
 XX  
 OS Homo sapiens.  
 OS Gallus gallus.  
 OS Synthetic.  
 OS Chimeric.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..18 /note= "Chicken heparanase signal peptide"  
 FT Region 19..527 /note= "Corresponds to residues 35-543 of human  
 FT heparanase mutant E225A (SEQ ID NO:7)"  
 FT Misc-difference 209 /note= "Ala replaces wild-type Glu (active site proton  
 FT donor). Corresponds to residue 225 of human heparanase  
 FT mutant E225A (SEQ ID NO:7)"  
 FT Active-site 327 /note= "Active site nucleophile"  
 FT  
 XX WO2004048558-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 24-NOV-2003; 2003WO-IL000989.  
 XX  
 PR 24-NOV-2002; 2002IL-00153059.

XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
XX PI Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
XX DR WPI; 2004-450373/42.  
XX DR N-PSDB; ADO63819.  
XX New nucleic acid construct comprising heparanase-derived polypeptide,  
PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
PT thrombasthenia, or Bernard-Soulier syndrome.  
XX PS Claim 10; SEQ ID NO 10; 128pp; English.  
XX The invention relates to nucleic acid constructs comprising a nucleic  
CC acid encoding a heparanase-derived protein which lacks heparanase  
CC endoglycosidase catalytic activity but which retains its cell-cell and  
CC cell-matrix adhesion properties. The constructs of the invention  
CC optionally further comprise operably linked regulatory elements. The  
CC invention also relates to the heparanase-derived proteins and host cells  
CC comprising the nucleic acid constructs of the invention. The heparanase-  
CC derived proteins are especially mutants of human heparanase in which the  
CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
CC host cells are useful in preparing a tissue sealant composition for  
CC sealing injuries, reducing the loss of blood, accelerating the healing  
CC and homeostasis of an injury, accelerating blood vessel endothelium  
CC formation or the endothelialisation of vascular grafts, accelerating the  
CC adhesive activity of mammalian cells, and accelerating the adhesion and  
CC aggregation of platelets. They may also be used in the treatment of  
CC disorders associated with adhesion deficiency such as LAD (leukocyte  
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
CC cells of the invention may additionally be used to screen for modulators of  
CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
CC synthetic vascular graft comprising a tube made of a biocompatible  
CC material lined with the cells. The present sequence represents a chimeric  
CC protein comprising the signal peptide of chicken heparanase and residues  
CC 35-543 of the human heparanase mutant E225A.  
XX SQ Sequence 527 AA;

Query Match 62.7%; Score 1711; DB 8; Length 527;  
Best Local Similarity 62.9%; Pred. No. 3.7e-164;  
Matches 332; Conservative 83; Mismatches 107; Indels 6; Gaps 3;

QY 1 MLVLLLLVLLAVPPRTA---ELQLGREPIGAVSPFLSLTLDASLARDPRFVALLR 56  
DB 1 MLVLLLLVLLAVPPRTAQDVLDLFFQEPHLVSPFLSVTIDANLATDPRFLILG 60  
QY 57 HPKLLTASGLSPGFLRFGTSTDFLIFPNKDSWEEKVLSFQA-KVCEAWPSFAV 115  
DB 61 SPKRLTLARGLSPAYLRFQGTGKDFLIFDPKSTFEERSYQSNQDICKYGIPTDV 120  
QY 116 PKLLLTQWLPLOEKLLIAEHSWKHKVTTITRSLDILHTFPASSGFLVFGIALLRRAG 175  
DB 121 BEKRLLEWYQEQLLREHYQFKNSTYSSRSSVDVLYTFANGSGLDLIFGLNALLRTAD 180  
QY 176 LOWDSSNAKOLLGYCAORSYNIWELGNENPSFRKSGICIDGFLGRDFVHLRQLLSQH 235  
DB 181 LQWNSNAQLLDLYCSKGNYSWELGNAPNSFLKADIFINGSQGEDFIQHLKLL-RK 239  
QY 236 PLYRHAELYGLDVGPQRKHTQHLLRFSKGGKAIDSVTWHYVNGRSATREDFLSPEV 295  
DB 240 STFKNAKLGPDVGPQPRKTKAKMLKSLKAGGEVIDSVTHYVNGRTATREDFLNPDV 299  
QY 296 LQSFATAIHDLVIGVATVPGKVKWLGEGTSAGVGGGAPQLSNYTVAGFMWLDKGLAARR 355  
DB 300 LDIFTSSQVKVQFQVSTPRGKVKWLGEGTSAGVGGGAPQLSDTFAAGFMWLDKGLSARM 359  
QY 356 GIDVVMRQVSFGAGSYHLVDAGPKPLPDYWLSSLKYLKLVGTRVLOASVEQADARRPRVYL 415

DB 360 GIEVVMRQVFFGAGNYHLVDENFDPLPDYWLSSLFKLVGTGKVLMAVSGQSKRRKLRVYL 419  
QY 416 HCTNPRPKYREGDVTLPALNLSNVTQSLQLPKQWLSKSDVQYLLPHGKDSILSREVQL 475  
DB 420 HCTNTDNPYKREGDLTLYAINLHNVTYLRPLPFPFSNKKQVDKYLRLPLGPHGLLSKSVQL 479  
QY 476 NGRLLQWVDDETLPALHEMALAPGSLGLPAFSYGFVYVIRNAKAIACI 523  
DB 480 NGLTLKWVDQDTLPPLMEKPLRFOSSGLGLPAFSYFFVIRNAKVAACI 527  
RESULT 10  
ID ADO63826  
AD ADO63826 standard; protein; 527 AA.  
XX AC ADO63826;  
XX DT 26-AUG-2004 (first entry)  
XX DE Chimeric heparanase mutant E343A, SEQ ID:11.  
XX KW Human; chicken; heparanase; heparanase-derived protein;  
KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;  
KW matrix adhesion; tissue sealant; injury; blood loss; endothelialisation;  
KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;  
KW adhesion disorder; LAD; leukocyte adhesion deficiency; drug screening;  
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome;  
KW vulnery; mutant; mutein.  
XX OS Homo sapiens.  
OS Gallus gallus.  
OS Synthetic.  
OS Chimeric.  
XX FH Key Location/Qualifiers  
FT Peptide 1..18  
FT Region /note= "Chicken heparanase signal peptide"  
FT /note= "Corresponds to residues 35-543 of human  
FT heparanase mutant E343A (SEQ ID NO:8)"  
FT Active-site 209  
FT /note= "Active site proton donor"  
FT Misc-difference 327  
FT /note= "Ala replaces wild-type Glu (active site  
FT nucleophile). Corresponds to residue 343 of human  
FT heparanase mutant E343A (SEQ ID NO:8)"  
XX WO2004048558-A2.  
XX 10-JUN-2004.  
XX 24-NOV-2003; 2003WO-IL000989.  
XX 24-NOV-2002; 2002IL-00153059.  
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
XX Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
XX WPI; 2004-450373/42.  
XX N-PSDB; ADO63820.  
XX New nucleic acid construct comprising heparanase-derived polypeptide,  
PT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
PT thrombasthenia, or Bernard-Soulier syndrome.  
XX PS Claim 10; SEQ ID NO 11; 128pp; English.  
XX The invention relates to nucleic acid constructs comprising a nucleic  
CC acid encoding a heparanase-derived protein which lacks heparanase  
CC endoglycosidase catalytic activity but which retains its cell-cell and  
CC cell-matrix adhesion properties. The constructs of the invention  
CC optionally further comprise operably linked regulatory elements. The  
CC invention also relates to the heparanase-derived proteins and host cells  
CC comprising the nucleic acid constructs of the invention. The heparanase-  
CC derived proteins are especially mutants of human heparanase in which the  
CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
CC host cells are useful in preparing a tissue sealant composition for  
CC sealing injuries, reducing the loss of blood, accelerating the healing  
CC and homeostasis of an injury, accelerating blood vessel endothelium  
CC formation or the endothelialisation of vascular grafts, accelerating the  
CC adhesive activity of mammalian cells, and accelerating the adhesion and  
CC aggregation of platelets. They may also be used in the treatment of  
CC disorders associated with adhesion deficiency such as LAD (leukocyte  
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
CC cells of the invention may additionally be used to screen for modulators of  
CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
CC synthetic vascular graft comprising a tube made of a biocompatible  
CC material lined with the cells. The present sequence represents a chimeric  
CC protein comprising the signal peptide of chicken heparanase and residues  
CC 35-543 of the human heparanase mutant E225A.

CC optionally further comprise operably linked regulatory elements. The  
CC invention also relates to the heparanase-derived proteins and host cells  
CC comprising the nucleic acid constructs of the invention. The heparanase-  
CC derived proteins are especially mutants of human heparanase in which the  
CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
CC host cells are useful in preparing a tissue sealant composition for  
CC sealing injuries, reducing the loss of blood, accelerating the healing  
CC and homeostasis of an injury, accelerating blood vessel endothelium  
CC formation or the endothelialisation of vascular grafts, accelerating the  
CC adhesive activity of mammalian cells, and accelerating the adhesion and  
CC aggregation of platelets. They may also be used in the treatment of  
CC disorders associated with adhesion deficiency such as LAD (leukocyte  
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet  
CC function), or Bernard-Soulier syndrome (deficient platelet adhesion). The  
CC cells of the invention may additionally be used to screen for modulators of  
CC cell-cell and cell-matrix adhesion, and to prepare an implantable  
CC synthetic vascular graft comprising a tube made of a biocompatible  
CC material lined with the cells. The present sequence represents a chimeric  
CC protein comprising the signal peptide of chicken heparanase and residues  
CC 35-543 of the human heparanase mutant E343A.  
XX  
SQ Sequence 527 AA;

Query Match 62.7%; Score 1711; DB 8; Length 527;  
Best Local Similarity 62.9%; Pred. No. 3.7e-164;  
Matches 332; Conservative 83; Mismatches 107; Indels 6; Gaps 3;  
QY 1 MVLVLLVLLVAVPPRTA---ELQGLREPIGAVSPFLSLTLDASLADPRFVALLR 56  
DB 1 MVLVLLVLLVAVPPRTAQDVVDLDFTEQPLHLVSPFLSVTIDANLATDPRFLILG 60  
QY 57 HPKLTLASGLSGFLRFGCTSTDFLFPNKSTWEEKVLSFQA-KDVCEAWPSFAV 115  
DB 61 SPKLTARGLSPAYLRFPGTKTDFLFPDKKSTFEERSYWSQVNDICKYSGIPPDV 120  
QY 116 PKLLTQWPLQEKLLLAHESKWKHNTTITRSTLDILHTFASSGFLRFLGNALLRRAG 175  
DB 121 EEKRLWEPQEQQLLREHQKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLTAD 180  
QY 176 LOWDSSNAKOLLYCAORSYNIWELGNPNRFRKSGICIDGFLQGRDFVHLRQLLSQH 235  
DB 181 LQWSSNAQLLDYSSKGYNIWELGNPNRFLKADIFINGSQLGEDFQLHKL-LRK 239  
QY 236 PLYRHAELGDLVGQPKRHTQHLRSPMKSGGKAIDSVTHHYVNGRSATREDFLSPEV 295  
DB 240 STEFNAKLYGPDVGQPRKRTAKMLKSLFKAGGEVIDSVTHHYVNGRTATREDFLNPDV 299  
QY 296 LDSFATAIHDLVIGVEATVPKVKWLGETSGAYGGAPQLSNTYVAGFMWLDKGLAARR 355  
DB 300 LDIFSSVQKVFQVVESTREPKVWLGTSAVGGAPLSDTPAAGFMWLDKGLSARM 359  
QY 356 GIDVWRQVSFGAGSYHLVDAGFKPLPDYWLSLYLKRLVGRVYLQASVEQADARRPVYL 415  
DB 360 GIEVWRQVFFGAGNYHLVDENPDYWLSSLFVKLVGTKVLMAVQSGKRLRVYL 419  
QY 416 HCTNPRHPKYREGDVTLPALNLGNVTQSLQPKOLWMSKVDQYLLPHGKDSILSRVQL 475  
DB 420 HCTNTDNPRYKEGDLTYALNLHNVTKYLRPLPYFPFSNKQVDKYLLRPLGPHGLLSKSVQL 479  
QY 476 NGRLLQWVDETLPALHEMALPGSTLGLPAFSGVFFVIRNAKAIACI 523  
DB 480 NGLTLKMVDQTLPLMEKFLRPGSSGLPFAFSYFFVIRNAKVAACI 527

RESULT 11  
ADO63827  
ID ADO63827 standard; protein; 527 AA.  
XX  
AC ADO63827;  
XX

DT 26-AUG-2004 (first entry)  
XX Chimeric heparanase mutant E225A/E343A, SEQ ID:12.  
DE  
XX Human; chicken; heparanase; heparanase-derived protein;  
KW heparanase mutant; non-catalytic; enzymatically inactive; cell adhesion;  
KW matrix adhesion; tissue graft; injury; blood loss; endothelialisation;  
KW blood vessel; vascular graft; platelet adhesion; platelet aggregation;  
KW adhesion disorder; LAD; leukocyte adhesion deficiency;  
KW Glanzmann's thrombasthenia; Bernard-Soulier syndrome; drug screening;  
XX vulnery; mutant; mutein.  
OS Homo sapiens.  
OS Gallus gallus.  
OS Synthetic.  
OS Chimeric.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..18  
FT /note= "Chicken heparanase signal peptide"  
FT Region 19..527  
FT /note= "Corresponds to residues 35-543 of human  
FT heparanase mutant E225A/E343A (SEQ ID NO:9)"  
FT Misc-difference 209  
FT /note= "Ala replaces wild-type Glu (active site proton  
FT donor). Corresponds to residue 225 of human heparanase  
FT mutant E225A/E343A (SEQ ID NO:9)"  
FT Misc-difference 327  
FT /note= "Ala replaces wild-type Glu (active site  
FT nucleophile). Corresponds to residue 343 of human  
FT heparanase mutant E225A/E343A (SEQ ID NO:9)"  
XX  
PN WO2004048558-A2.  
XX  
PD 10-JUN-2004.  
XX  
PF 24-NOV-2003; 2003WO-IL000989.  
XX  
PR 24-NOV-2002; 2002IL-00153059.  
XX  
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
XX  
PI Vlodavsky I, Zecharia E, Goldshmidt O, Ilan N;  
XX WPI; 2004-450373/42.  
DR N-PSDB; ADO63821.  
XX  
XX New nucleic acid construct comprising heparanase-derived polypeptide,  
FT useful in treating wounds, Leukocyte Adhesion Deficiency, Glanzmann's  
FT thrombasthenia, or Bernard-Soulier syndrome.  
PS  
PS Claim 10; SEQ ID NO 12; 128pp; English.  
CC  
CC The invention relates to nucleic acid constructs comprising a nucleic  
CC acid encoding a heparanase-derived protein which lacks heparanase  
CC endoglycosidase catalytic activity but which retains its cell-cell and  
CC cell-matrix adhesion properties. The constructs of the invention  
CC optionally further comprise operably linked regulatory elements. The  
CC invention also relates to the heparanase-derived proteins and host cells  
CC comprising the nucleic acid constructs of the invention. The heparanase-  
CC derived proteins are especially mutants of human heparanase in which the  
CC active site proton donor Glu225 and/or the active site nucleophile Glu343  
CC are replaced with Ala (ADO63822-ADO63824), and the proteins may  
CC optionally further comprise an avian heparanase signal peptide (ADO63825-  
CC ADO63827). The heparanase-derived protein, nucleic acid construct and  
CC host cells are useful in preparing a tissue sealant composition for  
CC sealing injuries, reducing the loss of blood, accelerating the healing  
CC and homeostasis of an injury, accelerating blood vessel endothelium  
CC formation or the endothelialisation of vascular grafts, accelerating the  
CC adhesive activity of mammalian cells, and accelerating the adhesion and  
CC aggregation of platelets. They may also be used in the treatment of  
CC disorders associated with adhesion deficiency such as LAD (leukocyte  
CC adhesion deficiency), Glanzmann's thrombasthenia (defective platelet







Db 133 PPVVEKRLLEWYQEQQLLREHYQKFKNSTYSSVDVLYTPANCSDGLDIFLGNALL 192  
QY 172 RRAGLQWSSNAQKLLGYCAQRSYNIISWELGNEPNSFRKSGICIDGQFGLRDFVHLRQL 231  
Db 193 RTADLQWSSNAQKLLDYSKSGYNIISWELGNEPNSFLKXADIFNGSLGEDFIQLHKL 252  
QY 232 LSQHLRYHAELVGLDVGPQRKHTQHLRLSRFMSKGGKKAIDSVTWHYVNGRSATREDFL 291  
Db 253 L-RKSTFKNAKLYGPDVGPQRKHTQHLRLSRFMSKGGKKAIDSVTWHYVNGRSATREDFL 311  
QY 292 SPEVLDSFATAIHDLVGLIVEATVPGKVMVGETGSAYGCGAPOLSNYVAGFWMLDKLGL 351  
Db 312 NPVDLIDIFISSVQKVFQVVESTRPGKVMVGETGSAYGCGAPOLSNYVAGFWMLDKLGL 371  
QY 352 AARRGIDVVMRQVSFGAGSYHLVDAGFKPLPDYWLSSLYKRLVTRVLOASVEQADARRP 411  
Db 372 SARMGIEVVMRQVFFGAGNYHLVDANFDPLPDYWLSSLYKRLVTRVLOASVEQADARRP 431  
QY 412 RYVLHCTNPRHPKYRGDVTLPALNLSNVTQSLQLPKQMSKVSVDQYLLPLPHGKDSILSR 471  
Db 432 RYVLHCTNPRHPKYRGDVTLPALNLSNVTQSLQLPKQMSKVSVDQYLLPLPHGKDSILSR 491  
QY 472 EVOLNGLRLQWDDQTLPLMEKPLRPGSSGLGPAFSYFFVIRNAKAAIACI 523  
Db 492 SVQNLGLRLQWDDQTLPLMEKPLRPGSSGLGPAFSYFFVIRNAKAAIACI 543

## RESULT 13

ABP56822  
ID ABP56822 standard; protein; 545 AA.  
AC ABP56822;  
DT 02-APR-2003 (first entry)  
XX Human heparanase protein SEQ ID NO:18.  
DE Human; heparanase; phosphorothioate; antisense oligonucleotide;  
KW cytosolic; gene therapy; tumour; enzyme.  
KW Homo sapiens.  
OS WO2003004705-A1.  
PN 16-JAN-2003.  
PD 01-JUL-2002; 2002WO-US020636.  
XX 05-JUL-2001; 2001US-00899440.  
PR (UYCO ) UNIV COLUMBIA NEW YORK.  
PA Stein C;  
PI WPI; 2003-201558/19.  
DR N-PSDB; AB222816.  
XX New oligonucleotide having a sequence complementary to a sequence of  
PT ribonucleic acid encoding a heparanase, useful for preparing a  
PT composition for treating tumor.  
XX Disclosure; Page 46-47; 48pp; English.  
PS The present invention describes an oligonucleotide having a sequence  
XX complementary to a sequence of ribonucleic acid encoding a heparanase.  
CC The oligonucleotide hybridises with the ribonucleic acid under conditions  
CC of high stringency and has a sequence comprising 10-40 bp. The  
CC internucleoside linkages of the oligonucleotide comprise at least one  
CC phosphorothioate linkage. Hybridisation of the oligonucleotide to the  
CC ribonucleic acid inhibits expression of the heparanase, where inhibition  
CC of heparanase means at least a 50% reduction in the quality of  
CC heparanase. Also described: (1) a method of inhibiting expression of a

CC heparanase in a cell; (2) a composition comprising the above  
CC oligonucleotide in an amount effective to inhibit the expression of  
CC heparanase in the cell and a carrier; and (3) a method of treating a  
CC tumour in a subject comprising administering to the subject an amount of  
CC the above oligonucleotide effective to inhibit expression of a heparanase  
CC in the subject. Heparanase antisense oligonucleotides have cytostatic  
CC activity, can be used in gene therapy, and can be used for preparing a  
CC composition for treating tumours. The present sequence represents human  
CC heparanase, which is given in the exemplification of the present  
XX invention

## SQ Sequence 545 AA;

Query Match 60.5%; Score 1650; DB 6; Length 545;  
Best Local Similarity 60.1%; Pred. No. 6.3e-158;  
Matches 321; Conservative 87; Mismatches 114; Indels 12; Gaps 3;  
QY 1 MLVLLLVLLAVPP-----RRTAELQLGRLPIGAVSPAFSLTLDASLARDPR 50  
Db 13 LMLLLGLPLGPGALPRPAQAQDVVDLDFQEPHLVSPFLSVTIDANLATDPR 72  
QY 51 FVALLRHPKLTASGLSPGFLRFGSTGDTLIFNPKDSTWEEKVLSFQA-KDYCEAM 109  
Db 73 FLILGSPKLTARGLSPAVLRFGGTTDFLFPKXSTFEERSYMQSQVNODICKY 132  
QY 110 PSFAVVPKLLTQWPLQEKLLAHSWKHKNTTITRSTLDLHTFASSSGFLVFLNA 169  
Db 133 SIPPVVEKRLLEWYQEQQLLREHYQKFKNSTYSSVDVLYTPANCSDGLDIFLGN 192  
QY 170 LLRAGLQWSSNAQKLLGYCAQRSYNIISWELGNEPNSFRKSGICIDGQFGLRDFVHLR 229  
Db 193 LLRTADLQWSSNAQKLLDYSKSGYNIISWELGNEPNSFLKXADIFNGSLGEDFIQLH 252  
QY 230 QLLSQHLRYHAELVGLDVGPQRKHTQHLRLSRFMSKGGKKAIDSVTWHYVNGRSATRED 289  
Db 253 KLL-RKSTFKNAKLYGPDVGPQRKHTQHLRLSRFMSKGGKKAIDSVTWHYVNGRSATRED 311  
QY 290 FLSPVLDSFATAIHDLVGLIVEATVPGKVMVGETGSAYGCGAPOLSNYVAGFWMLDKL 349  
Db 312 FLNPDVLDIFISSVQKVFQVVESTRPGKVMVGETGSAYGCGAPOLSNYVAGFWMLDKL 371  
QY 350 GLAARRGIDVVMRQVSFGAGSYHLVDAGFKPLPDYWLSSLYKRLVTRVLOASVEQADAR 409  
Db 372 GLSARMGIEVVMRQVFFGAGNYHLVDENFDPLPDYWLSSLYKRLVTRVLOASVEQADAR 431  
QY 410 RPYVLHCTNPRHPKYRGDVTLPALNLSNVTQSLQLPKQMSKVSVDQYLLPLPHGKDSIL 469  
Db 432 KRLVHCTNPRHPKYRGDVTLPALNLSNVTQSLQLPKQMSKVSVDQYLLPLPHGKDSIL 491  
QY 470 SREYQLNGLRLQWDDQTLPLMEKPLRPGSSGLGPAFSYFFVIRNAKAAIACI 523  
Db 492 SKSVQNLGLRLQWDDQTLPLMEKPLRPGSSGLGPAFSYFFVIRNAKAAIACI 545

## RESULT 14

ADE16012  
ID ADE16012 standard; protein; 545 AA.

XX AC ADE16012;

XX DT 29-JAN-2004 (first entry)

XX DE G-coupled protein receptor related polypeptide, SEQ ID No 42.

XX G-coupled protein receptor; antidiabetic; anorectic; antibacterial;  
KW virucide; fungicide; cytostatic; nootropic; neuroprotective;  
KW antiparkinsonian; haemostatic; antileptemic; neurogenesis;  
KW cell differentiation; cell proliferation; hematopoiesis; wound healing;  
KW angiogenesis; gene therapy; chromosome mapping; tissue typing;  
KW preventive medicine; pharmacogenomics; human.

XX Homo sapiens.

XX

PN WO200283841-A2.  
XX 24-OCT-2002.  
XX 03-APR-2002; 2002WO-US010713.  
XX 03-APR-2001; 2001US-0281136P.  
XX 05-APR-2001; 2001US-0281863P.  
XX 05-APR-2001; 2001US-0281906P.  
XX 10-APR-2001; 2001US-0282934P.  
XX 13-APR-2001; 2001US-0283657P.  
XX 13-APR-2001; 2001US-0283678P.  
XX 13-APR-2001; 2001US-0283687P.  
XX 13-APR-2001; 2001US-0283710P.  
XX 17-APR-2001; 2001US-0284234P.  
XX 19-APR-2001; 2001US-0285325P.  
XX 20-APR-2001; 2001US-0285609P.  
XX 23-APR-2001; 2001US-0285748P.  
XX 23-APR-2001; 2001US-0285890P.  
XX 24-APR-2001; 2001US-0286068P.  
XX 27-APR-2001; 2001US-0287213P.  
XX 03-MAY-2001; 2001US-0288509P.  
XX 30-MAY-2001; 2001US-0294495P.  
XX 31-MAY-2001; 2001US-0294801P.  
XX 31-JUL-2001; 2001US-0309216P.  
XX 25-SEP-2001; 2001US-0324775P.  
XX 28-NOV-2001; 2001US-0333900P.  
XX 02-APR-2002; 2002US-00115479.  
XX (CURA-) CURAGEN CORP.  
XX Li L, Gerlach V, Liu X, Miller CE, Spytek KA, Zerhusen BD;  
PI Pena CE, Shenoy SG, Zhong H, Smithson G, Caaman SJ, Boldog FL;  
PI Voss EZ, Vernet CM, Macdougall JR, Rastelli L, Anderson DW;  
PI Zhong M, Mezes PD, Furtek K, Paturajan M, Burgess CE, Malyankar UM;  
PI Shinkets RA, Taupier RJ, Edinger SR, Mazur A;  
XX WPI; 2003-067574/06.  
XX N-PSDB; ADL61011.  
XX New isolated NOVX polypeptides and polynucleotides, useful for  
PT preventing, diagnosing or treating NOVX-associated disorders e.g.  
PT diabetes, obesity, dyslipidemias, cancer, Parkinson's disease,  
PT Alzheimer's disease, infections.  
XX Claim 1; SEQ ID NO 42; 320pp; English.  
XX The invention relates to a novel isolated G-coupled protein receptor  
CC related polypeptides. The novel polypeptide comprise any of the 22 fully  
CC defined sequences of 87-1780 amino acids, given in the specification;  
CC their mature forms; and possible variants. The novel polypeptides have  
CC the following activities: antidiabetic, anorectic, antibacterial,  
CC virucide, fungicide, cytostatic, and antilipaeamic. The G-coupled protein  
CC antiparkinsonian, haemostatic, and antilipaeamic. The G-coupled protein  
CC receptor related polypeptides are useful in a method of treating or  
CC preventing in a human, a pathology associated with the G-coupled protein  
CC receptor related polypeptides. The polypeptides are useful in the  
CC manufacture of a medicament for treating a syndrome associated with a  
CC human disease, preferably a NOVX-associated disorder. The novel  
CC polypeptides are useful for treating, preventing or diagnosing diseases,  
CC such as metabolic disorders, diabetes, obesity, infectious diseases,  
CC anorexia, cancer-associated diseases, neurodegenerative disorders,  
CC Alzheimer's disease, Parkinson's disease, immune disorders, hematopoietic  
CC disorders, and various dyslipidemias, metabolic disturbances associated  
CC with obesity, metabolic X syndrome and wasting disorders associated with  
CC chronic diseases and various cancers. The nucleic acids and polypeptides  
CC may also be used as targets for the identification of small molecules  
CC that modulate or inhibit e.g. neurogenesis, cell differentiation, cell  
CC proliferation, hematopoiesis, wound healing and angiogenesis, in gene  
CC therapy, in generation of antibodies that bind immunospecifically to NOVX  
CC substances for use in therapeutic or diagnostic methods. The nucleic  
CC acids are further used as hybridization probes, in chromosome mapping,  
CC tissue typing, preventive medicine, and pharmacogenomics. This sequence

CC represents one of the novel G-coupled protein receptor related  
CC polypeptides of the invention.  
XX  
SQ Sequence 545 AA;  
Query Match 60.5%; Score 1650; DB 7; Length 545;  
Best Local Similarity 60.1%; Pred. No. 6.3e-156;  
Matches 321; Conservative 87; Mismatches 114; Indels 12; Gaps 3;  
QY 1 MLVLLLVLLAVPP-----RRTAELQLGREPIGAVSPAFSLTLDASLARDP 50  
DB 13 LLMLLLLGPIGPSFALPRPAQAQDVLDFTFQEPHLVSPFLSVTTIDANLATDP 72  
QY 51 FVALLRHPKLTSLASGLSPGFLREGGTTDFLI FPNKNDSTWEEKVSEFOA-KDVCEAW 109  
DB 73 FLILGSPKLTTLARGLSPAYLREGGTTDFLI FDPKKESTFEERSYQSQVNQDICKYG 132  
QY 110 PSFAVPKLLLTQWPLQEKLLLAHSHWKKHNTTITRSTLIDLTFPASSSSFLRVFGLNA 169  
DB 133 SIPPDVEEKLEWPFYQELLREHYQKFKNSTYSRSSVDVLYTFANCGLDILFGLNA 192  
QY 170 LLRRAGLOWSSNAQLLGYCAORSYNI SWELGNEPNSFKKSGICIDGFGOLGRDFVHLR 229  
DB 193 LLRTADLQWSSNAQLLDLYCSSKGINISWELGNEPNSFLKKADIFINGSQLGEDFIQLH 252  
QY 230 QLLSQHPLYRHAELYGLDVGPQRKHTQHLLRSFMKSGKKAIDSVTWHYYVNGRSATRED 289  
DB 253 KLL-RKSTFNKLYGPDVGPQRKTA KLSFLKAGEVIDSVTWHYYLNGRTATRED 311  
QY 290 FLSPEVLDSFATAIHDLVIGIVEATVPCKKWLGTSGSAYGGAPOLSNYYVAGFMWLDKL 349  
DB 312 FLNPDVLDIFISSVQKVFQVVESTRPCKKWLGTSGSAYGGAPOLSNYYVAGFMWLDKL 371  
QY 350 GLAARGIDVVMQVSFGAGSYHLVDAGFKPLPDYWLISLLYKRLVGTFLQASVEQADAR 409  
DB 372 GLSARMGIEVVMQVFFGAGNYHLVDENFDPLPDYWLISLLYKRLVGTFLQASVEQADAR 431  
QY 410 RPRVYLCTNPRHPKYREGDVTLPALNLSNVTSQSLQPKQLWSKSDVDQILPHGKDSIL 469  
DB 432 KLRVYLCTNPRHPKYREGDVTLPALNLSNVTSQSLQPKQLWSKSDVDQILPHGKDSIL 491  
QY 470 SREVQLNGRLLQWDDDTLPALHEMALAPGSTLGLPAPSYGYFVIRNAKAIACI 523  
DB 492 SKSVQLNGLTLKMWDDDTLPALMEKPLRPGSSSLGLPAPSYGYFVIRNAKVAACI 545

## RESULT 15

ADL93951  
ID ADL93951 standard; protein; 545 AA.

AC ADL93951;

DT 20-MAY-2004 (first entry)

DE Human G-coupled protein receptor-related protein #21.

XX human; transgenic; Gene Therapy; Protein Therapy; cardiomyopathy;  
XX atherosclerosis; hypertension; congenital heart defect; aortic stenosis;  
XX atrial septal defect; atrioventricular canal defect; ductus arteriosus;  
XX pulmonary stenosis; subaortic stenosis; ventricular septal defect;  
XX valve disease; tuberosclerosis; scleroderma; obesity; transplantation;  
XX adrenoleukodystrophy; congenital adrenal hyperplasia; prostate cancer;  
XX neoplasm; adenocarcinoma; lymphoma; uterus cancer; fertility;  
XX haemophilia; hypercoagulation; idiopathic thrombocytopenic purpura;  
XX immunodeficiency; graft versus host disease; AIDS; bronchial asthma;  
XX Crohn's disease; G-coupled protein receptor; metabolic disorder;  
XX neurodegenerative disorder; receptor.

OS Homo sapiens.

XX US2004006205-A1.

XX 08-JAN-2004.

PD



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:10:07 ; Search time 17.4126 seconds  
(without alignments)  
2889.939 Million cell updates/sec

Title: US-10-645-659A-5  
Perfect score: 2728  
Sequence: 1 MLVLLLVLLAVPPRTAE.....LPAPSYGYVIRNAKAIACI 523

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: Piri:\*  
2: Pir2:\*  
3: Pir3:\*  
4: Pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	810.5	29.7	480	JC7506	heparanase protein
2	395	14.5	521	T45608	hypothetical prote
3	128.5	4.7	190	T01953	hypothetical prote
4	108.5	4.0	1004	S53939	probable membrane
5	106.5	3.9	1225	C84330	hypothetical prote
6	106	3.9	503	A48546	genome polyprotein
7	105	3.8	473	S71934	genome polyprotein
8	105	3.8	496	GNX31E	genome polyprotein
9	104.5	3.8	374	B82488	hypothetical prote
10	102.5	3.8	1180	A11939	two-component hybr
11	101	3.7	1076	A69409	carbamoyl-phosphat
12	99	3.6	987	D97029	ribonucleotide red
13	98.5	3.6	836	D64988	yejO protein - Esc
14	97.5	3.6	284	JS0652	aminoglycoside N3'
15	97.5	3.6	780	S44560	alpha, alpha-trehal
16	97.5	3.6	836	A98014	hypothetical prote
17	97.5	3.6	836	C85858	hypothetical prote
18	97	3.6	341	A57136	class I histocompa
19	97	3.6	493	JC1327	protective antigen
20	96.5	3.5	565	A10479	probable membrane
21	96.5	3.5	627	C98148	acetoin catabolism
22	96.5	3.5	627	A13139	transcription regu
23	95.5	3.5	1012	GNX3AU	genome polyprotein
24	95	3.5	4568	T08030	dynamin beta heavy
25	94.5	3.5	395	C84138	8-amino-7-oxononan
26	94	3.4	584	AH2506	ser/thr protein ki
27	94	3.4	867	D87320	conserved hypothet
28	94	3.4	897	G02529	dynamin heavy chain
29	93.5	3.4	485	T07596	1-aminocyclopropan

30	93.5	3.4	492	2	F70326	conserved hypothet
31	93.5	3.4	500	2	D87541	beta-xylosidase [i
32	93.5	3.4	567	2	T44363	poly(3-hydroxybuty
33	93	3.4	401	2	B83377	probable phospholi
34	93	3.4	515	2	AE0186	probable decarboxy
35	93	3.4	1039	2	C87083	C-term lysyl-tRNA
36	92.5	3.4	341	2	JC5663	major histocompati
37	92	3.4	893	2	T15183	hypothetical prote
38	92	3.4	1839	1	OYBYK	adenylate cyclase
39	92	3.4	2108	2	H70819	probable polyketid
40	91.5	3.4	454	2	JC4616	apyrase (EC 3.6.1.
41	91.5	3.4	964	2	S45944	hypothetical prote
42	91.5	3.4	1012	1	GNXS52	genome polyprotein
43	91.5	3.4	1379	2	T13718	pollux gene protei
44	91	3.3	884	2	H83322	hypothetical prote
45	91	3.3	911	2	A56465	transcription fact

ALIGNMENTS

RESULT 1

JC7506

heparanase protein 2a - human

C:Species: Homo sapiens (man)

C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 09-Jul-2004

C:Accession: JC7506

R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hircock, M.; Pat

Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000

A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family me

A:Reference number: JC7506

A:Accession: JC7506

A:Molecule type: mRNA

A:Residues: 1-480 <MCK>

A:Cross-references: UNIPROT:Q9HB39; UNIPARC:UPI000003E88A; GB:AF282885

A:Comment: This protein, a intracellular membrane-bound enzyme, has biological and therai

C:Genetics:

A:Gene: hpa2a

A:Map position: 10q23-10q24

C:Keywords: heparin binding; membrane bound

Query Match 29.7%; Score 810.5; DB 2; Length 480;  
Best Local Similarity 34.8%; Pred. No. 5.1e-58;  
Matches 192; Conservative 69; Mismatches 161; Indels 129; Gaps 9;

QY	1	MLVLLLVLLA-----VPPRTAEIQ-----LGLREPIGAVSPAFSLTLTLDAS	44
DB	27	LYLALLHLSSQNGDRRPLVDRAAGLKEKTLILLDVTSTKNPVTNVNFIQLQDPS	86
QY	45	LARDPRFVALLRHPKLTLAGSLSPGLRFGTSTDFLIF-----NPNKDSWEEKVLSEF	100
DB	87	IIHD-GWLDLSSKRLVTLARGLSPAFLRFGGKRTDFLQFQNLNPAKS-----	134
QY	101	QAKDVCEAWPFAVVPKLLLTQWPLQELKLLAEHSKKHKNVTITRSLDILHTFPASSG	160
DB	135	-----	134
QY	161	FRLVFGLNALLRRAGLQWDSNAKQLLGYCAQRSYNISWELGNEPNSFRKSGICIDGFO	220
DB	135	-----RGPGPD-----YYLKNYE--DEPNNYTMGVRANGSQ	166
QY	221	LGRDFVHLRQLLSOPLRYHAELGYDVGQPKRHTQHLLRSFMKSGKAIDSVTWHYYV	280
DB	167	LGDYIQLKSLQPIRIYRSRSLYGNIGRPNKVNIALLDGPMKVGASTVDVATWQHCVI	226
QY	281	NGRSATREDFLSPVLDLSFAITHDVLGIVEATVPGKKVWLGETSGAYCGGAPQLSNTYV	340
DB	227	DGRVVKVMDFLKTRLLDLSQIRKIQKVNTYTPGKKIWLGVVTTTSAGGTNNLSDSYA	286
QY	341	AGFWLWDLGLAARGIDVVMRQVSGAGSVHLVDAGFKPLPDYWLISLLYKRLVGRVLQ	400
DB	287	AGFLWNLTLGLANQGIIDVIRHSFFDHGYNHLVDQNFNPLPDYWLISLLYKRLIGPKVLA	346

Qy	401	ASVQADAR-PP-----RVYLHCTPRHPKPYREGDVTLPALNLSNVTQSLQPLKQLW	451
		:	
Db	347	VHVAGLQKRPGRVIRDKLRIYAHCTHHNHNYYRGSITLFIINLHRSRKKIKLAGTLR	406
		:	
Qy	452	SKSDVOYLLPHGKDSILSRVOLNGRLLQWDDDTLPALHEMALAPGSLGLPAFSVGF	511
		:	
Db	407	DKLVHGYLLQPYGQEGELKSKSVQLNGQPLVMVDDGTLPBLKPRPLRAGRRTLVPVPTMGF	466
		:	
Qy	512	YVIRNAKAIAIC	522
		:	
Db	467	FVVKGNVALAC	477
		:	

RESULT 2  
 T45608  
 hypothetical protein F13G24.30 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 09-Jul-2004  
 R:Accession: T45608  
 R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckae:  
 submitted to the Protein Sequence Database, December 1999  
 A:Reference number: 223009  
 A:Accession: T45608  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-521 <BEV>  
 A:Cross-references: UNIPROT:Q9SDA1; UNIPARC:UPI00000A497C; EMBL:AL133421  
 A:Experimental source: cultivar Columbia; BAC clone F13G24  
 C:Genetics:  
 A:Map position: 5  
 A:Introns: 53/3; 66/1; 127/2; 177/1; 256/3; 319/2; 361/2; 394/3  
 A:Note: F13G24.30

**RESULT 3**

T01953  
hypothetical protein T2L5.6 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 09-Jul-2004  
C:Accession: T01953  
R;Geisel, C.; Smith, A.; Le, T.  
submitted to the EMBL Data Library, October 1998  
A:Description: The sequence of A. thaliana T2L5.  
A:Reference number: Z14470  
A:Accession: T01953  
A>Status: translated from GB/EMBL/DDBJ  
Molecule type: DNA  
A:Molecule type: DNA  
A:Residues: 1-190 <GEI>  
A:Cross-references: UNIPROT:O82604; UNIPARC:UIP00000A8F7D; EMBL:AF096371; NID:g3695386; PID:  
A:Experimental source: cultivar Columbia  
C:Genetics:

A:Map position: 4  
A:Introns: 36/2; 69/3  
A>Note: T2L5.6  
C:Superfamily: Arabidopsis thaliana hypothetical protein T2L5.6

Query Match                4.7%; Score 128.5; DB 2; Length 190;  
Best Local Similarity     25.1%; Pred.No. 0.0052;  
Matches          47; Conservative      35; Mismatches      70; Indels      35; Gaps      8;

QY            362 RVSFGAGSYHLVD-AGFKPLPDYWLSLLYLKRLVGTFRVLQASVEQADARRPRVVLTCTNP    420  
Db            ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
                 12 QSLIG-GNYGLVTNTTTPDPDYYSALIWRQLMGKALFTTF--SGTKIRSYTHCAR-    67  
  
QY            421 RHPKYREGDVTLFALNLSNVQT-----SLOLPKL---WSKVSDGYDLLPHG---    464  
Db            :    :    :    :    :    :    :    :    :    :    :    :    :  
                 68 -----QSKGITVLMNLNDTTTVAKVELNNFSLSRHTHKMSYKRASSQLFCGPNGVIQ    122  
  
QY            465 -----KSIL-SREVOLNRLQMVDDETLPALHEHALAPGSTLGLPAPSYGFFVIR    515  
Db            |    :    :    :    :    :    :    :    :    :    :    :    :    :  
                 123 REEYHLTAKDGNIHSQTMLLNQNALQVNSMGDLPPIETIHINSTEPTIAPYSIVVFVHR    182  
  
QY            516 NKAIAIC 522  
Db            |    ||  
                 183 NVVPAC 189

**RESULT 4**

S53939  
Probable membrane protein YGL241w - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein G0733; hypothetical protein HRC1004  
C:Species: Saccharomyces cerevisiae  
C>Date: 08-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 09-Jul-2004  
C:Accession: S53939; S60489; S64266  
R;Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.  
submitted to the EMBL Data Library, April 1995  
A:Description: The sequence of a 11.1 kb DNA fragment between ADH4 and ADE5 on the left arm of  
A:Reference number: S53934  
A:Accession: S53939  
Molecule type: DNA  
A:Molecule type: DNA  
A:Residues: 1-1004 <VAN>  
A:Cross-references: UNIPROT:P53067; UNIPARC:UIP0000052FA2; EMBL:Z49149; NID:g793865; PID:  
A:Experimental source: strain S288C  
R;Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.  
Yeast 11, 1519-1523, 1995  
A>Title: The sequence of an 11.1 kb DNA fragment between ADH4 and ADE5 on the left arm of  
A:Reference number: S60484; MUID:96353434; PMID:8750240  
A:Accession: S60489  
A>Status: nucleic acid sequence not shown; translation not shown  
Molecule type: DNA  
A:Molecule type: DNA  
A:Residues: 1-1004 <VAN>  
A:Cross-references: UNIPARC:UIP0000052FA2; EMBL:Z49149; NID:g793865; PID:NCAA89014.1; PI:  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995  
R;Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.  
submitted to the Protein Sequence Database, May 1996

A:Reference number: S64263  
A:Accession: S64266  
A:Molecule type: DNA  
A:Residues: 1-1004 <VAF>  
A:Cross-references: UNIPARC:UPI0000052FA2; EMBL:Z72763; NID:g1322906; PID:e243899; PID:G  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:KAP114  
A:Cross-references: SGD:S0003210  
A:Map position: 7L  
C:Keywords: transmembrane protein  
F:186-202/Domain: transmembrane #status predicted <TMM>

Query Match 4.0%; Score 108.5; DB 2; Length 1004;  
Best Local Similarity 20.8%; Pred. No. 2.6;  
Matches 120; Conservative 84; Mismatches 212; Indels 161; Gaps 29;

QY 16 RRTAEQLQ--GLREPIGAVSPFLSLTLDASLARDPRFVALLRHPKLTLASGLSGFLR 73  
DB 18 REVAETQLQWDSQASQVFKALANVALQHEASLESQFALLSLRLKITWY--WSPGFES 75  
QY 74 FGGTST-----DF-----LIFNPKNKSTWBEKV--LSEFOAKDVCEAMPS-FAVVP 116  
DB 76 YRSTNSVEIDVKDFIREVLKLCINDNENTKINGASYCIVQISAVDFPDWQQLTVIY 135  
QY 117 KLLLTOWPLOEKLLAE-----HSWKHKHNTVITRSTDLHTFTASSSGRLVFGNAL- 170  
DB 136 DAISHQSHLSNMSLLNEIYDDVVSSEMFPEGGIGLATMEIVFKVLNTETSTLIAKAAK 195  
QY 171 LRAGL-----QWSSNAKQLLGYCAQBSYNIWEL-----GNPNSPRKSGCIDG 218  
DB 196 LLKACLLQWSSHNEYDEASRKSFSVQCLATSLQILQQLTLNFGN-----VD- 242  
QY 219 FOLGRDFVHLRQLLSOHLRYHAELVGLDVQPKKH-TOHLRSFMKSGGAIDSVTWHH 277  
DB 243 -----VISQLKFKSIYENLVFIKNDPS--RKFFSSELQKFIMALQDLENVT--- 289  
QY 278 YYNGRSATRE-----DFLT-----SPEVLDSFATAIHDVLGIVEAT 313  
DB 290 -HINANVTETSESPLELTVHDCSIYIPELTSVCTLQFSVEEMNKIITSLTILCOLSSET 348  
QY 314 VPQKKWLGETSGAYGGAPQLSNVYVAGFWMLDKLGLAARRGIDVVMRQVSGAGSYHL 373  
DB 349 ---REIWTSD-----FNTFVS-----KETGLAA-----SYNV 372  
QY 374 VDAG--FKPLPDYWLKLVGRVRLQVQAEQADARRPRVYLHCTNPRHPKYREGDV 430  
DB 373 RDQANEFTSLPNPQLSLPK-FVSDIEHSTCNSTLESLLYLQCI-----LIMNDEI 426  
QY 431 TLPALNSNVQSLQLPKQLWSKSDYOYLLPHGKDSILSRVOLGRLLQMWDDTLPA 490  
DB 427 T-----GENIDQSL-----QILIKTENILVSOEIPELLARAILTIPRVLDKDI-ALPD 476  
QY 491 LHEMA---LAPGSTGLP-----APSYGFY 512  
DB 477 IKPLTSAFLAKSLNALKSKDKELIKSATLIAFTYYCY 513

RESULT 5  
C84530  
Hypothetical protein At2g15540 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: C84530  
R.;lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
eusa, D.; Niemman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: C84530  
A:Status: preliminary  
A:Molecule type: DNA

```

A;Residues: 1-1225 <STO>
A;Cross-references: UNIPROT:Q9ZQF8; UNIPARC:UPI00000A75F2; GB:AE002093; NID:g4335720; PFI
C;Genetics:
A;Gene: At2g15540
A;Map position: 2

Query Match      3.9%; Score 106.5; DB 2; Length 1225;
Best Local Similarity 20.9%; Pred. No. 5.1;
Matches 126; Conservative 86; Mismatches 233; Indels 159; Gaps 34;

Qy      10  LLAVPPRTAELQGLREPTG--AVSPAFLSLTLDASLARDPRFVALLRH-----PKLH 61
Db      356  LQEVPLVITEMANKSLTKVISPVEVKRALFSLNPDKARPGDGTAFYQHYHDLTGPDLI 415

Qy      62  TLASGL-SPGLRFGGTSDFLFPNPKDSTWEEKVLSEFOAKDVCEAWPSPVAPKL- 119
Db      416  KLVQNFHSTGFFEDRLNETNICLI-----PKTERPKMAEFRPISLCNV--SYKVISKVL 469

Qy      120  -----LQWPLQEKLLLAHSHWKKHKNYTTITRSTLDLHTPASSGFF 161
Db      470  SRLKRLPELISETQSAFVAERLITDNLIAQENFHALRTNPACCKYKMAIKTDSKAYD 529

Qy      162  RLVRG-LNALLRAG-----LQWSSNAKQLLGYC-AQRSYNIWELGNEPNSFRKSGI 214
Db      530  RVESFELRALMLKMGFAQKWDW-----IIFCISVSYKIL--LNGSPKGFIKPS-- 577

Qy      215  CIDGFQGRDFVHLRQLLSQHPLYR-----HAELYGLDVQPRKHTOHLRS-----F 262
Db      578  --RGIRQGPISPLFLFILTALVAKLKDAEMHGRIQGLQISRASPSHTLLFADDSLFF 635

Qy      263  MKSG---GKAIDSVTHHHYVNGRSATREDPLSEVLDSFATAIHDVLGIVEATVPGKV 319
Db      636  CKADPLQKBEIIDRLRYGEASGOQ-----LNPD--KSSVMFGHEVDNSINTI---KV 684

Qy      320  WLGETGSAYGGA--POLSNITYVAGFMWLDKGLAARRGI-----DVVMRQVSFGAGSYH 372
Db      685  SLG-----IHKDGAIRSKLSSV-VANFWKTR---EESNGIHWIAWDKLCITPFDGG--- 732

Qy      373  LVDAGFKPLPDYMLSLLYKRL-----VGTRVLQA-----SVEQADARRPRVYLH 416
Db      733  ---LGFRTLEEFNLVLLAKQLWRLIRFPNSLLSRVLRGRYFRYSDDPIQIGKANRPSFGWR 789

Qy      417  CTNPRHPKYREGDVTLPALNLSNVQSL--QLPKQLWKSVDQYL--LLPHGKDSILS-R 471
Db      790  SIMAACP-----LLLSGLRRTIGSGLMTRVME--DPWIPSPFPPRAKSIILNR 835

Qy      472  EVOLN-----GRLLQWVDETLPALHEMALAPGSTLGLPAFSYGR-----Y 512
Db      836  DTHLYVNDLIDPVTKQWKGLRQLQELVDPDSIPLI--LGIRPSRTYKSDDFSWSFTKSGNY 893

Qy      513  VIRN 516
Db      894  TVKS 897

RESULT 6
A48546
genome polyprotein - infectious bursal disease virus (strain A) (fragment)
C;Species: infectious bursal disease virus
C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 31-Dec-2004
C;Accession: A48546
R;Lana, D.P.; Beisel, C.E.; Silva, R.F.
Virus Genes 6, 247-259, 1992
A;Title: Genetic mechanisms of antigenic variation in infectious bursal disease virus: ar
A;Cross-references: UNIPROT:Q04094; UNIPARC:UPI00001786F9
A;Note: sequence extracted from NCBI backbone (NCBIN:114640, NCBIIP:114641)
C;Genetics:
A;Accession: A48546
A;Molecule type: genomic RNA
A;Residues: 1-503 <LAN>
A;Cross-references: UNIPROT:Q04094; UNIPARC:UPI00001786F9
A;Note: sequence extracted from NCBI backbone (NCBIN:114640, NCBIIP:114641)
C;Genetics:
A;Map position: segment A
C;Keywords: polyprotein

```

Query Match	3.9%;	Score 106;	DB 2;	Length 503;
Best Local Similarity	24.1%;	Pred. No. 1.5;		
Matches	69;	Conservative 46;	Mismatches 109;	Indels 62; Gaps 16;

  

Qy	232	LSOHPLYRHAELCYLDVGQPRKHTQHLLRFMKSGGKAI-----DSVTWHHYV--NG	282
Db	33	LEKHTURSETSYNLTVGD-----TGSLIVFPFGPGSIGVCAHYTLQSNQ	78
Qy	283	RSATREDFLSPEVL-----SPATAIHDVLGIVEATVPGKKVWLGETSA--YGGGAPOLSN	337
Db	79	NYKFDQMLLTAQNLPPASYNCRLVRSRLTVRSSTLPGGVYALNGTINAVTFQGSLSBELD	138
Qy	338	TYVAGFMWL-----DKLG-LAARRGIDVVMRQVSFCAGSYHLVDAGFKPLDPYWLISLYK	391
Db	139	VSYNGLMSATANINDKIGNVLVEGVTLSLPTSVDLGYVRLGD----PIPAICLD---P	191
Qy	392	RLVGTVRVLQASVDAQADRRPRVYLHCNTNPH-----PKYREGDV--TLFALNLSNVTTQSLQL	446
Db	192	KWV-----ATCDDSD--RPRVVTITAADYQFSQYQCGGVITILFSAIDAIT--SLSV	242
Qy	447	PKQLWSKSVDQYLLLP-----GKD--SILSREVQLNGRLLQWDD	485
Db	243	GGSELVFKTSVQSIVLAGATIYLIQFGDGTAVITRAVAANGLTAGIND	288

RESULT 7

S71934

genome polyprotein - infectious bursal disease virus (strain E/DEL) (fragment)

N;Contains: major structural protein VP2; nonstructural protein VP4

C;Species: infectious bursal disease virus

A;Variety: strain E/DEL

A;Note: host Gallus gallus (chicken)

C;date: 19-Mar-1997 #sequence\_revision 29-Aug-1997 #text\_change 09-Jul-2004

C;Accession: S71934; S60291

R;Vakharia, V.N.

submitted to the EMBL Data Library, October 1990

A;Reference number: S71934

A;Accession: S71934

A;Molecule type: genomic RNA

A;Residues: 1-473 <YAK>

A;Cross-references: UNIPROT:Q64957; UNIPARC:UPI00000F6949; EMBL:X54858; NID:g486622; PII

A;Experimental source: strain E/DEL

R;Vakharia, V.N.; Ahmed, B.; He, J.

Avian Dis. 36, 736-742, 1992

A;Title: Use of polymerase chain reaction for efficient cloning of dsRNA segments of inf

A;Reference number: S60291; MUID:93038354; PMID:1329714

A;Accession: S60291

A;Residues: 206-350 <VMA>  
A;Cross-references: UNIPARC:UPI00001786F6; EMBL:X54858  
A;Experimental source: strain E/DL  
C;Comment: This virus is responsible for a severe immunodepressive disease in young chicks  
C;Genetics:  
A;Map position: segment A  
C;Superfamily: birnavirus  
C;Keywords: nonstructural protein; polyprotein; structural protein  
F;1-453/Product: major structural protein VP2 #status predicted <VP2>  
F;453-4473/Product: nonstructural protein VP4 (fragment) #status predicted <VP4>

	Query Match	3.8%;	Score 105;	DB 2;	Length 473;
	Best Local Similarity	24.1%;	Pred. NO. 1.6;		
	Matches	69;	Conservative	45;	Mismatches 110; Indels 62; Gaps 16;
Qy	232	LSOHPHYRHAELYGLDVGQPKTKTQHLLRSFMSGGKAI-----DSVTWHHHYYV--NG	282		
	:				
Dd	33	LKHTRLRSETSYNLTVGD-----TGSLIVFPFFPGFSIVGAHYTTQQSNG	78		
Qy	283	RSATREDFLSPVELD---SFATAIHVDVLGIVEATVPCKKWILGETGSA--YGCAGAPQLSN	337		
	:   :   :   :   :   :   :   :   :   :   :				
Dd	79	NYKFQDMLLTAQNLPASINYCRVLVSRLTVRSRTPLFGVVVALNGTINAVTFQGSLELTD	138		
Qy	338	TYVAGFMWL-----DKLG-LAARRRGTDVMVRQYSFGAGSYHLDVAGFKPLDPDWLSLLYK	391		

```

139 VSYNGLMSATANIINDKIGNVLGEGVTVLSLPTSVDLGVTVRLGD-----PIPATGLD---P 191
1392 RLVGTVRLVQASVEQADARRPVYLHCTNPRH---PKYREGDV--TLFALNLSNVTQSLQL 446
192 KMW-----ATCDSSD--RPRVYITTAADNYQFSQYQTGGVTVITLFSANIDAIT-SLSV 242
447 PKQLWSKSDVQYLLPH-----GKD--SILSREVQLNGRLLQWVD 485
243 GGELVFTKSVQSLVGATYILIGFDGTAVITRAVAANGLTAGIDN 288

```

RESULT 8  
GNXSIE  
genome polypeptide - infectious bursal disease virus (strain E) (fragment)  
N/Contains: major structural protein VP2; nonstructural protein VP4  
C/Species: infectious bursal disease virus  
A/Note: host Gallus gallus (chicken)  
C/Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 09-Jul-2004  
C/Accession: FQ0283  
R/Reine, H.G.; Haritou, M.; Failla, P.; Fahey, K.; Azad, A.  
J. Gen. Virol. 72, 1835-1843, 1991  
A/Title: Sequence analysis and expression of the host-protective immunogen VP2 of a variant  
A/Reference number: FQ0283; PMID:91341469; PMID:1651980

A;Molecule type: Genomic RNA  
A;Residues: 1-496 <HEI>  
A;Cross-references: UNIPROT:P29802; UNIPARC:UPI00000615FC; GB:D10065; NID:G221889; PIDN:P29802  
C;Comment: This virus is responsible for a severe immunodepressive disease in young chicks  
C;Genetics:  
A;Map position: segment A  
C;Superfamily: birnavirus  
C;Keywords: nonstructural protein; polypeptide; structural protein  
F;1-452/Product: structural protein VP2 #status predicted <VP2>  
F;453-496/Product: nonstructural protein VP4 (fragment) #status predicted <VP4>

Query Match 3.8%; Score 105; DB 1; Length 496;  
Best Local Similarity 24.1%; Pred. No. 1.8;  
Matches 69; Conservative 45; Mismatches 110; Indels 62; Gaps 16;

Qy	232	LSQHPLYRHAELYGLDVGQPKKHTQHLRLSPWKSCKAI-----DSVTWHHYV--NG 282
	:	:
	:	:
Db	33	LEKHTRLRSETSYNLNTGVD-----TGSGLIIVFPFGPGSGTVGAHYYLQSNG 78

```

283  RSATREDFLSPVLVD---SFATAIHDVLGI VEATVPKKVWLGETGSA--YGGGAQLSN 337
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
79  NYAFDOMLLTAQNLPAASNYCRLLVRSRTIVRASSTLPGGVYALNGTINAVTFQGSGLGELTD 138

```

QY            338   TYVAGFWL-----DKLG-LAARRGIDVVMRQVSFGAGSYHLVDAGFKPLPDYWLSQLYK   391

D6            139   VSNGLMASATINDKIGNVLVGSGVTLSLPTSDYLGVYRGLD-----PIPAIGLD---P   191

QY 392 RLVGTRVLQASVEQADARRPRVYLHCTNPRH---PKYREGDV--TLFALNLSNVTOSLQL 446

nb 192 KMV-----ATCQSSD--RPRVYITADNVPQSSCYOTGGVITLFSANIDAIT--SLSV 242

447 PKOLWSKVQYLPH-----GKD--SILSREVQLNGRLLQWVDD 485

RESULT 9  
B82488  
hypothetical protein VCA0200 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
C;Species: Vibrio cholerae  
C;date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 31-Dec-2004  
C;Accession: B82488

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M. Nature 406, 477-483, 2000

A:Title: DNA Sequence of the cholera pathogen *Vibrio cholerae*.  
A:Reference number: A82035; MUID:20408833; PMID:10952301



Qy	112	FAVVPKLLTQWPLOEQKLLLAESHWSKKHKNVTITRSTLDILH-----TFASSSGSFRLL 163
Db	482	LAVAELK-----STEITATDLDVVNEQIIDKVKNHQSKGQK- 518
Qy	164	VFGNALRLRAGLQWDSNNAQLLGYCAQRSYNIISWELGNEPSPKKGICIGDGFOLGR 223
Db	519	--GSSRVVEVEGLERINVLASELLIHQRRTLNDE-----QVQE 555
Qy	224	DFVHLRQLLSQHPHYRHAELVGLDVGQPRKHTQHLLRSFMKSGGKAIDSVTWHYYVNGR 283
Db	556	IIIEQLQLRNLKHQETLN-QLRDLPLQORSLATOH-TQSPASVNFDPLEMDAYTEFHLTLH 613
Qy	284	SATREDFLSPVILDSF-----ATAIHD-----VLGIVEATVPGKKVWLGETSGSAYGGGA 333
Db	614	EAIETSQLOQATESIDFLFQATQIISHKKQSLAFSLDNLVEARMPLPLENILSRPFHMV 673
Qy	333	PQLSNTVYAGFMWMDKLGLAARRGIDVVMRQVSFGAGSVHLVDAGF-KPLPDYWLISLYK 393
Db	674	KNLSTVY-----HKRVDLKP-----TGTVDVLKATAEKLDPILLHLV-- 711
Qy	392	RLVGRVRLQASVEQADARRPR-----VYLHCTNPRHPKYRGEDVTLFALNLSNVTQSLQ 445
Db	712	----RNAFDHGIESPVRERKGPQGVIEICA-----YHQSQTI--LEVRDDGGGLK 759
Qy	446	LPKQLMSKSVDOYLLPHGKDSILSRVOLN-----GRLLQVMDVETLPALHEMAL 496
Db	760	L-----DSIRKKAIELNFISKKDESKGYLSQINDSELL-----EFMF 796
Qy	497	APGSTLIG-----LPAFSYGVFYVRN 516
Db	797	APGFTTAEKVSSELGRGMGLDIVRS 821

RESULT 11

A69409

carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) large chain C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004

C:Accession: A69409

R;Klenk, H.P.; Clayton, R.A.; Tomb, J.P.; White, O.; Nelson, K.E.; Ketchum, K. ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkland, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L. Nature 390, 364-370, 1997

A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing bacterium Pyrococcus furiosus

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: A69409

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Cross-references: 1-1076 <KLE>

C:Function:

A:Pathway: glutamate metabolism; pyrimidine nucleotide biosynthesis

C:Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain C:Keywords: ligase; pyrimidine nucleotide biosynthesis

F:9-1055/Domain: carbamoyl-phosphate synthase (Glutamine-hydrolyzing) large chain F:9-469/Domain: biotin carboxylase homology <BC1>

F:562-994/Domain: biotin carboxylase homology <BC2>

	Query Match	3.7%	Score 101;	DB 2;	Length 1076;
	Best Local Similarity	20.7%	Pred. No. 12;		
	Matches	75;	Conservative 49;	Mismatches 136;	Indels 102; Gaps 14
Qy	186	LLGYCAQRSYNISWELGNEPNSFRKSGCIGCIDGFGQDRDFVHLRQLLSQHPLYRHAELYG	245		
Db	640	IVQFGGQTPLNIAREL-----EDSGARILGTSVDSDIAEDRRERFAELLER-----	685		
Qy	246	LDVGQRKRKHTOH-----LLRSFMKSGGKAI-----DSVTWHHYVYNGRSA	285		
Db	686	LNIDPPENGIAHSLEAAKEIARKTGFPVLRPSYVLGGRMELIVDDEETLERIVITEALEV	745		
Ov	286	TREDFLSPEVLDSPTAIDHVLGI-VEATVPQKVKWLG-----ETGSAYGGGAPQLSN	337		

Db 746 SPE---KPIIIDKF---LEDAIEVEVDALCDGEVWVIGGIMEHIEEAGVHSGDSA----- 794  
QY 338 TYVAGFWMLDKGLAARGIDVVMRQVSGAGSYHLVDAGFKPLDPLDWLSL-LYKRLVGT 396  
Db 795 -----CVLPVPS-----LDEVTINTIVDTYKRLALANVVL 826  
QY 397 RVULOASVEQADARRPRVYLHCTNPRHPKYREGDVTLPALNLSNVTQSLQPLKQLWSKSV 456  
Db 827 INIQYAVKDG-----KYVLEANPRASRTVPFVSKATGIPLAKIAAKLMMGKKLRELGVK 881  
QY 457 QYLLLPH-----GKDSILSRVQLNGRLLQWDDETLPALHEMALAPGSTL 502  
Db 882 EKULKHVAVKEAVFPPIKLPGVDPVLGPENKSTGEVNGIDYDFGL-AYYKAELAAAGMKL 940  
QY 503 GL 504  
Db 941 PL 942

RESULT 12  
D97029  
ribonucleotide reductase, vitamin B12-dependent [imported] - Clostridium acetobutylicum  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C;Accession: D97029  
; R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld  
A;Accession: D97029  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-987 <KUR>  
A;Cross-references: UNIPROT:Q97K72; UNIPARC:UPI000000CA083; GB:AE001437; PIDN:AAK79023.1;  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC1047

Query Match 3.6%; Score 99; DB 2; Length 987;  
Best Local Similarity 18.2%; Pred. No. 15;  
Matches 65; Conservative 52; Mismatches 112; Indels 128; Gaps 14;

QY 125 LQSKLLAEHSWKHKNTTITRSTLDLHTFASSS-GFRLVFGNLALLRRAGLQWDSSNA 183  
Db 442 LVQLLEASHWGQFPTEDEVART-----HMFRTIGLQ---ISNLAALVMVNGPYDSEEG 494

QY 184 K-----QLLGYCAQRSYNISWELGNENPNSFRKKSIGICIDGFGQGRDFVHLRQLLSQHPLXR 239  
Db 495 RAISSSLIGVLTGYSYISSLMAKEVGAPEK----- 525

QY 240 HAEYGLDVQPKRHTQHLRLSRPMKSGGK---AIDSVTHHHYVNGRSATRED--FLSPE 294  
Db 526 ----YNIN---KEHMLKVLNRSARAAGSIDTPEFKIGYKPLVNVNHSILEKEDLRYSSE 577

QY 295 VLDSFATAIHDLVGIVEATVPGKKVLGFTGSAYGAGGAPOLSNVTYVAGFWMLDKLGLAAR 354  
Db 578 LKNSMNSAL-----ESGEKYGFRNAQVSNAPVG----- 606

QY 355 RGIDVVMRQVSGAGSYHLVDAGFKPLDPLDWLSLKYRLVGT RVULOASVEQADARRPRVY 414  
Db 607 -----TISFA-----MDCASTSVEPFSSHVIYKLSGGG-----Y 636

QY 415 LHCTNP-----RHPKYREGDVTLPALNLSNVTQSLQPLKQLWSKSVDOYLLPHGK 465  
Db 637 MTLTNPLIQESLHGLGYSENEI-----EDITNYILRKKKVKTDDCGTEYETILDGK 686

RESULT 13  
D64988  
yeJ0 protein - Escherichia coli (strain K-12)  
C;Species: Escherichia coli

C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C;Accession: D64988  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Coi  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: A64720; MUID:97426617; PMID:9278503  
A;Accession: D64988  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-836 <BLAT>  
A;Cross-references: UNIPARC:UPI0000168114; GB:AE000308; GB:U00096; NID:g1788508; PIDN:AA  
A;Experimental source: strain K-12, substrain MGL655  
C;Genetics:  
A;Gene: yeJ0

Query Match 3.6%; Score 98.5; DB 2; Length 836;  
Best Local Similarity 22.5%; Pred. No. 13;  
Matches 72; Conservative 36; Mismatches 109; Indels 103; Gaps 17;

QY 134 HSWKKHKN-TTITRSTLDLHTFASSSGFRVLVFGNLALLRRAGLQWDSSNAKQLIGYCAQ 192  
Db 523 HSWSLAENRAQIPTSITDVLNMAAQP---LVFDA-----ELDTVRERLGSVKG 568

QY 193 RSYNIS-WEIG-NEPNSFRKKSGI---CIDGFGQGRDFVHLRQLLSQHPLRYHAEYGL 246  
Db 569 VSYDTAMWSAINTRNNVTVDAGAGFEOTLTGLTGLID-----SRFSREESSTIRGL 620

QY 247 DVGQPKRHTQHLRLSRPMKSGKAIDSVT-----WHHY---YVNGRSATREDFLSPEVLD 297  
Db 621 IFGY-----SHSDIGFDRGKGKGNIDSYTLGAYAGMEHQNGAYVDG-----VWKVD 665

QY 298 SFATAIHDLVGIVEATVPGKKVWLGETGSAYGAGGAPOLSNVTY-AGFWMLDKL----- 349  
Db 666 RFANTH-----GK---MSNGATAFGDYNNGAGAHVSGFRWVDGLSVRPYL 711

QY 350 -----GLAARRGIDVVMRQVSGAGSYHLVDAGFKPLDPLDWLSLKYRL 393  
Db 712 AFTGFTTGDQDYTLNSGMRADVGNTRILRAEAGTAVSYHM-DLQNGTTLEPWLKAAVQK- 769

QY 394 VGT RVULOASVEQADARRPRV 413  
Db 770 -----EYADSNQVKV 779

RESULT 14  
JS0652

aminoglycoside N3'-acetyltransferase (EC 2.3.1.81) - Streptomyces griseus (strain SS-119)  
C;Species: Streptomyces griseus  
C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004  
C;Accession: JS0652

R;Ishikawa, J.; Hotta, K.

Gene 108, 127-132, 1991

A;Title: Nucleotide sequence and transcriptional start point of the kan gene encoding an  
A;Reference number: JS0652; MUID:92104494; PMID:1761222

A;Accession: JS0652

A;Molecule type: DNA

A;Residues: 1-284 <ISH>

A;Cross-references: UNIPROT:Q54216; UNIPARC:UPI000000B0D8B; GB:D00681; NID:g217013; PIDN:I  
C;Comment: This enzyme confers kanamycin resistance.

C;Genetics:

A;Gene: kan

C;Superfamily: Escherichia coli aminoglycoside N3'-acetyltransferase

C;Keywords: acyltransferase; coenzyme A

Query Match 3.6%; Score 97.5; DB 2; Length 284;  
Best Local Similarity 26.8%; Pred. No. 3.2;  
Matches 57; Conservative 17; Mismatches 82; Indels 57; Gaps 12;

QY 304 HDVLGIVEATVPGKKVWLGETGSAYG--GGAPOLSNVTYVAGFWMLDKLGAARGIDVVM 361  
Db 21 HDLAAL--GLVPGDVTVMFHTRLSAIGYVSGGPQTV-----IDAL-----LDVV- 61

```
QY 362 ROVSFGAGSYHLVDAGFKPLPDY-----WLSLLYKRLVGTIVLQASVEQADARRPV 413
Db 62 -----GPTGTLVTCGNDAPPDYFTWPPAAQEAIVRAHHAFDPRTSEAEHANGRLPEA 116
QY 414 YLH---CTNPRHPKYREGDVTFLFALNLSNVTQSLQLPKQMSKSVDOYLLPLPHGKDSILS 470
Db 117 LRRRPGAVRSRHP-----DVSLLAALGAS-----APALMDAHPWDD-----PHGPGSPLA 160
QY 471 REVOLNORLLOM-VDETLPALHE---MALAPG 499
Db 161 RLVALGGRVLLGLGAPRDTMTLLHHAELAQAPG 193

RESULT 15
S44560
alpha,alpha-trehalase (EC 3.2.1.28) NTH2 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YBR001c; protein YBR0106
C;Species: Saccharomyces cerevisiae
C;Date: 08-Jun-1994 #sequence_revision 09-Sep-1994 #text_change 09-Jul-2004
C;Accession: S44560; S45852; S37322
R;Wolfe, K.H.; Lohan, A.J.E.
Yeast 10 (Suppl.A), S41-S46, 1994
A;Title: Sequence around the centromere of Saccharomyces cerevisiae chromosome II: similar
A;Reference number: S44556; MUID:94378721; PMID:8091860
A;Accession: S44560
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-780 <WOL>
A;Cross-references: UNIPROT:P35172; UNIPARC:UPI0000052EA3; EMBL:Z26494; NID:9403311; PID
R;Lohan, A.J.E.; Wolfe, K.H.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45730
A;Accession: S45852
A;Molecule type: DNA
A;Residues: 1-780 <LOH>
A;Cross-references: UNIPARC:UPI0000052EA3; EMBL:Z35870; NID:9536185; PIDN:CAA84937.1; PI
C;Genetics:
A;Gene: SGD:NTH2
A;Cross-references: SGD:S0000205; MIPS:YBR001c
A;Map position: 2R
C;Keywords: glycosidase; hydrolase

Query Match 3.6%; Score 97.5; DB 2; Length 780;
Best Local Similarity 22.0%; Pred. No. 14;
Matches 75; Conservative 49; Mismatches 110; Indels 107; Gaps 19;

QY 217 DGFQLGRDFVHLRQLLSQHL---YRHAELYGLDVGPQRKHTQ-HLLR-----S 261
Db 39 EGQGGRR--HRRLLSMHEYDFPSNAEVYGPITDPKQSKIHLRLNRTMTSVFNKVS 96
QY 262 FMKSGGKAIDSVTHHHYVNGRSATREDELSPE-----VLDSFATAIHVLG----- 308
Db 97 DFXNGMK-----DYTLKRGSEDDSFUSSQGNRRFYIDNVDLALDELLASEDTDKNH 149
QY 309 --IVEATVPKGVKWLGTSGAYGGAPQLSNVTYVAGFWMLDKGLAARRGIDVVMQVSF 366
Db 149 QITIEDT--GPKVIKVGTSANGFKNVNRGTMYLSNL-LQELTIK-----SF 194
QY 367 GAGSYHLVDAGFKPLP-----DYWLSL-----LYKRLVGTIVLQASVEQADARRP 411
Db 195 GRHQIFLDEARINENPVDRLSRLITTFWTSLTRRVDLYN--IAIARDSKIDTPGAKNP 252
QY 412 RVYL--HCTNPRHPKYREGDVTFLFALNLSNVTQSLQ-----LPKQMSKSVDOYLLPLPHG 464
Db 253 RIYVPYNC-----PEQYE-----FYIQASQMNPSLKLEVEYLPKIDITAEYVYKSLNDTP-- 300
QY 465 KDSILSREVQLNGVLLQWVDDETLPALHEMALPGSTIGLP 505
Db 301 -----GLLALAMEEHNPNSTGERSL-----VGYP 324
```

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:06 ; Search time 108.256 Seconds

(without alignments)  
2293.354 Million cell updates/sec

Title: US-10-645-659A-4

Perfect score: 2841

Sequence: 1 MLLRSKPALPPPLMLLLGP.....LPAPSYFFVIRNAKVAACI 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq 8:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*
- 9: Geneseq2005s:\*
- 10: Geneseq2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2841	100.0	543	2	AAY17082 Human hep
2	2841	100.0	543	4	AAB86206 Human hep
3	2841	100.0	543	7	ADD18950 Human dis
4	2841	100.0	543	8	ADK52086 Human ato
5	2841	100.0	543	8	ADM48759 Human hpa
6	2841	100.0	543	8	ADM05074 Antipsoxi
7	2841	100.0	543	8	ADN04902 Antipsoxi
8	2841	100.0	543	8	ADQ80372 Heparanas
9	2841	100.0	543	8	ADR88210 Human pre
10	2841	100.0	543	8	ADP25079 PRO poly
11	2841	100.0	543	8	ADT78177 Human hep
12	2841	100.0	543	9	ADY27036 Human hep
13	2841	100.0	543	9	AEA42426 Human hep
14	2841	100.0	588	2	AAY10124 A human p
15	2838	99.9	543	2	AAY02345 A human h
16	2838	99.9	543	3	AAY57590 Human hep
17	2838	99.9	543	3	AAB08849 Amino aci
18	2838	99.9	543	3	AAY52990 Human hep
19	2838	99.9	543	4	AAY97635 Human hep
20	2838	99.9	543	5	ABB07813 Human hep
21	2838	99.9	543	7	ADG88800 Human hpa
22	2838	99.9	543	8	ADL16379 Human hep
23	2838	99.9	543	8	ADM48716 Human hpa

24	2838	99.9	543	9	AEA42466	Aea42466 Human hep
25	2838	99.9	543	10	AE966848	Ae966848 Human hep
26	2838	99.9	592	2	AAY02346	Aay02346 A human h
27	2838	99.9	592	3	AAB08850	Aab08850 Amino aci
28	2838	99.9	592	7	ADG88804	Adg88804 Human SK-
29	2838	99.9	592	8	ADL16383	Adl16383 Human hep
30	2838	99.9	592	8	ADM48720	Adm48720 Human SK-
31	2838	99.9	592	9	AEA42461	Aea42461 Human hep
32	2835	99.8	543	8	ADO63831	Ado63831 Human hep
33	2835	99.8	543	8	ADO63823	Ado63823 Human hep
34	2835	99.8	543	8	ADO63832	Ado63832 Human hep
35	2835	99.8	543	8	ADO63822	Ado63822 Human hep
36	2829	99.6	543	4	AAB88361	Aab88361 Human mem
37	2829	99.6	543	8	ADO63824	Ado63824 Human hep
38	2829	99.6	543	9	ADY63087	Ady63087 Human clo
39	2824.5	99.4	556	9	ADZ19010	Adz19010 Heparanas
40	2820	99.3	545	6	ABP56822	Abp56822 Human hep
41	2820	99.3	545	7	ADE16012	Adel6012 G-coupled
42	2820	99.3	545	8	ADL93951	Adl93951 Human G-c
43	2807.5	98.8	570	9	ADZ19008	Adz19008 Heparanas
44	2767	97.4	530	2	AAY34173	Aay34173 Human pre
45	2740	96.4	532	2	AAY17083	Aay17083 Seq ID No

#### ALIGNMENTS

RESULT 1  
AAY17082  
ID AAY17082 standard; protein; 543 AA.  
XX  
AC AAY17082;  
XX  
DT 21-JUL-1999 (first entry)  
XX  
DB Human heparanase enzyme.  
XX  
KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;  
KW metastasis; angiogenesis; wound healing; angioplasty-induced restenosis;  
KW arteriosclerosis; atherosclerosis; inflammation; tissue development;  
KW human; HSPG.  
XX  
OS Homo sapiens.  
XX  
FN WO9921975-A1.  
XX  
PD 06-MAY-1999.  
XX  
PF 28-OCT-1998; 98WO-AU000898.  
XX  
PR 28-OCT-1997; 97AU-00000062.  
PR 09-DEC-1997; 97AU-00000812.  
XX  
PA (AUSU ) UNIV AUSTRALIAN NAT.  
XX  
PI Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;  
XX  
WPI; 1999-312956/26.  
XX  
N-PSDB; AAX37259.  
XX  
PT Polynucleotides encoding mammalian endoglucuronidases, especially  
PT heparanases, useful to promote wound healing.  
XX  
PS Claim 6; Page 69-73; 112pp; English.  
XX  
CC The invention relates to nucleic acid sequences that encode heparanase  
CC enzymes having endoglucuronidase activity. Recombinant heparanases are  
CC capable of removing the HS side chain from heparan sulfate proteoglycan  
CC (HSPG). Sulfated oligosaccharides, sulphates or HSPG can be used to  
CC inhibit heparanase, this is useful for treatment of a physiological or  
CC medical condition associated with elevated heparanase activity, such as  
CC metastasis, angiogenesis, wound healing, angioplasty-induced restenosis,  
CC arteriosclerosis, atherosclerosis and inflammation. The human, murine and

CC rat heparanases can be used to enhance wound healing, especially  
CC associated with tissue development and repair. The conditions mentioned  
CC above can be diagnosed using specific antibodies, and also using primers  
CC and probes specific for the heparanase polynucleotides. Other uses of the  
CC heparanases include sequencing sulfated molecules such as HSPG. The  
CC present sequence represents a human heparanase

XX  
SQ Sequence 543 AA;

Query Match 100.0%; Score 2841; DB 2; Length 543;  
Best Local Similarity 100.0%; Pred. No. 3.2e-275;  
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQAQDVVDLDFFTQBPFLHVSFSLSVT 60  
DB 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQAQDVVDLDFFTQBPFLHVSFSLSVT 60  
QY 61 IDANLATDPRFLILGSPKLTTLARGLSPAYLRFGGKTDFLIPDPKKESTFEERSYQWS 120  
DB 61 IDANLATDPRFLILGSPKLTTLARGLSPAYLRFGGKTDFLIPDPKKESTFEERSYQWS 120  
QY 121 QVNQDICKYGSIPDPVEEKLRLWPYQEOQLLRHYOKKFNSTYSRSSVDVLYTFANCS 180  
DB 121 QVNQDICKYGSIPDPVEEKLRLWPYQEOQLLRHYOKKFNSTYSRSSVDVLYTFANCS 180  
QY 181 GLDLIFGLNALLRTADLQWNSNAQALLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240  
DB 181 GLDLIFGLNALLRTADLQWNSNAQALLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240  
QY 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGGEVIDSVTWHYYL 300  
DB 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGGEVIDSVTWHYYL 300  
QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA 360  
DB 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA 360  
QY 361 AGFMWLDKLGSLARMGIEVVMRQVFFGAGNHYLVNDENFDPLDWLSLLFKKLVGTKVL 420  
DB 361 AGFMWLDKLGSLARMGIEVVMRQVFFGAGNHYLVNDENFDPLDWLSLLFKKLVGTKVL 420  
QY 421 ASVQGSRRKRLRVYLHCTNTDNPRYKEDLTLYAINLHNVTKYLRPLYPFSNKQVDKYL 480  
DB 421 ASVQGSRRKRLRVYLHCTNTDNPRYKEDLTLYAINLHNVTKYLRPLYPFSNKQVDKYL 480  
QY 481 RPLGPHGLLSKSVQLNGLTLKQVDDQTLPLMEKPLRPGSSLGLPAPFSYFFVIRNAKVA 540  
DB 481 RPLGPHGLLSKSVQLNGLTLKQVDDQTLPLMEKPLRPGSSLGLPAPFSYFFVIRNAKVA 540

RESULT 2

AAB86206  
ID AAB86206 standard; protein; 543 AA.

XX  
AC AAB86206;

XX  
DT 24-AUG-2001 (first entry)

XX  
DE Human heparanase inhibitor protein.

XX Heparanase; inhibitor; cardiac insufficiency; cardiast; nephrotropic;  
KW hepatocytic; veterinary medicine; congestive heart failure; dyspnoea;  
KW primary cardiomyopathy; peripheral odema; pulmonary congestion;  
KW hepatic congestion; hydrothorax; ascite; nocturia; human.

XX  
OS Homo sapiens.

XX  
PN DE19955803-A1.

XX  
↑

PD 23-MAY-2001.  
XX  
PF 19-NOV-1999; 99DE-01055803.  
XX  
PR 19-NOV-1999; 99DE-01055803.  
XX  
PA (KNOL ) KNOLL AG.  
XX  
PI Herr D, Hahn A, Laux V;  
XX  
DR WPI; 2001-368371/39.  
DR N-PSDB; AAH20940.  
XX  
PT Treatment or prevention of cardiac insufficiency and related conditions,  
PT e.g. pulmonary congestion and dyspnoea, comprises administration of  
PT heparanase inhibitor.  
XX  
PS Disclosure; Page 11-13; 16pp; German.  
XX  
CC This invention describes a novel heparanase inhibitor which can be used  
CC for the treatment or prevention of cardiac insufficiency and associated  
CC indications, symptoms and/or malfunctions. The heparanase inhibitor of  
CC the invention has cardiant, nephrotropic and hepatotropic activity. The  
CC products of the invention can be used in human and veterinary medicine,  
CC for the treatment or prevention of congestive heart failure e.g. primary  
CC cardiomyopathy. Associated conditions treated or prevented with the  
CC inhibitor are especially peripheral odemas, pulmonary and hepatic  
CC congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g.  
CC nocturia can also be treated. This sequence represents the human  
CC heparanase protein described in the method of the invention  
XX  
SQ Sequence 543 AA;

Query Match 100.0%; Score 2841; DB 4; Length 543;  
Best Local Similarity 100.0%; Pred. No. 3.2e-275;  
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQAQDVVDLDFFTQBPFLHVSFSLSVT 60  
DB 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQAQDVVDLDFFTQBPFLHVSFSLSVT 60  
QY 61 IDANLATDPRFLILGSPKLTTLARGLSPAYLRFGGKTDFLIPDPKKESTFEERSYQWS 120  
DB 61 IDANLATDPRFLILGSPKLTTLARGLSPAYLRFGGKTDFLIPDPKKESTFEERSYQWS 120  
QY 121 QVNQDICKYGSIPDPVEEKLRLWPYQEOQLLRHYOKKFNSTYSRSSVDVLYTFANCS 180  
DB 121 QVNQDICKYGSIPDPVEEKLRLWPYQEOQLLRHYOKKFNSTYSRSSVDVLYTFANCS 180  
QY 181 GLDLIFGLNALLRTADLQWNSNAQALLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240  
DB 181 GLDLIFGLNALLRTADLQWNSNAQALLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240  
QY 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGGEVIDSVTWHYYL 300  
DB 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGGEVIDSVTWHYYL 300  
QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA 360  
DB 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA 360  
QY 361 AGFMWLDKLGSLARMGIEVVMRQVFFGAGNHYLVNDENFDPLDWLSLLFKKLVGTKVL 420  
DB 361 AGFMWLDKLGSLARMGIEVVMRQVFFGAGNHYLVNDENFDPLDWLSLLFKKLVGTKVL 420  
QY 421 ASVQGSRRKRLRVYLHCTNTDNPRYKEDLTLYAINLHNVTKYLRPLYPFSNKQVDKYL 480  
DB 421 ASVQGSRRKRLRVYLHCTNTDNPRYKEDLTLYAINLHNVTKYLRPLYPFSNKQVDKYL 480  
QY 481 RPLGPHGLLSKSVQLNGLTLKQVDDQTLPLMEKPLRPGSSLGLPAPFSYFFVIRNAKVA 540  
DB 481 RPLGPHGLLSKSVQLNGLTLKQVDDQTLPLMEKPLRPGSSLGLPAPFSYFFVIRNAKVA 540

QY	541	ACI	543	61	IDANLATDPRFLILLGSPKLTARGLSPAYLRFGTKTDFLI	120
Db	541	ACI	543	61		120
QY	ADD18950			121	QVNOQICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNSTYSRSSVDVLYTFANC	180
Db	ADD18950 standard; protein; 543 AA.			121		180
AC	ADD18950;			181	GIDLI FGLNALLRTADLQWSSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS	240
XX				181		240
DT	15-JAN-2004 (first entry)			241	QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSFKAGGEVIDSVTWHHYL	300
DE	Human disease related protein SeqID439.			241		300
XX				301	NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKVKWLGETSSAYGGAPLLSDTFA	360
KW	human; disease state; cytostatic; antiinflammatory; ophthalmological;			301		360
KW	antiarteriosclerotic; vulnery; gene therapy;			361	AGFWMLDKGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLPDYMLSLFLKVLGVKVL	420
KW	hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;			361		420
KW	inflammation; erythropoiesis; glycolysis; gluconeogenesis;			421	ASVOGSKRRKLRVYLHCTNTDNPYKEGDLTLIYAINLHNVTKYLRPYFPFSNKQVDKYL	480
KW	glucose transportation; catecholamine synthesis; iron transport;			421		480
KW	nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;			481	RPLGPHGLLSKSVOLNGLTLKMWDDQTLPLPMEKPLRPGSSSLGLPAFSYSFFVIRNAKVA	540
KW	retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;			481		540
KW	inflammatory condition; wound healing.			541	ACI	543
OS	Homo sapiens.			541		543
XX				541	ACI	543
PN	WO2003018621-A2.					
XX						
PD	06-MAR-2003.					
XX						
XX	23-AUG-2002; 2002WO-GB003892.					
PF						
XX						
PR	23-AUG-2001; 2001GB-00020558.					
PR	05-OCT-2001; 2001GB-00024037.					
XX						
PA	(OXFO-) OXFORD BIOMEDICA UK LTD.					
XX						
PI	Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;					
XX						
DR	WPI; 2003-290046/28.					
DR	N-PSDB; ADD18951.					
XX						
PT	New substantially purified polypeptide, useful for diagnosing or treating					
PT	a hypoxia-regulated condition, such as cancer, ischemia, reperfusion					
PT	injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or					
PT	wound healing.					
XX						
PS	Claim 25; SEQ ID NO 439; 424pp; English.					
XX						
CC	This invention relates to novel human genes and gene product which are					
CC	implicated in certain disease states. Compounds which modulate the					
CC	proteins of the invention may have cytostatic, antiinflammatory,					
CC	ophthalmological, antiarteriosclerotic or vulnery activities. The					
CC	sequences of the invention may be useful for gene therapy. The invention					
CC	may be useful for diagnosing or treating a hypoxia-regulated condition,					
CC	such as tumorigenesis, angiogenesis, apoptosis, inflammation,					
CC	erythropoiesis, or the biological response to hypoxia conditions					
CC	including processes such as glycolysis, gluconeogenesis, glucose					
CC	transportation, catecholamine synthesis, iron transport or nitric oxide					
CC	synthesis. The disease includes cancer, ischaemic conditions, reperfusion					
CC	injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,					
CC	inflammatory conditions or wound healing. The present sequence is that of					
CC	a disease related protein of the invention.					
XX						
SQ	Sequence 543 AA;					
Query Match	100.0%; Score 2841; DB 7; Length 543;					
Best Local Similarity	100.0%; Pred. No. 3.2e-275;					
Matches 543; Conservative	0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MLLRSKPALPPPLMLLLGLPLSPGALPRPQAQDVLDLFTQBLHLVSPSFLSVT	60			
Db	1	MLLRSKPALPPPLMLLLGLPLSPGALPRPQAQDVLDLFTQBLHLVSPSFLSVT	60			

Detecting atopic dermatitis or psoriasis comprises assaying levels of expression of an indicator gene at a rash site and non-rash site of a

PT person with atopic dermatitis or psoriasis.  
XX Example 2; SEQ ID NO 119; 484pp; Japanese.  
XX  
CC The invention relates to detecting atopic dermatitis or psoriasis  
CC comprising assaying the levels of expression of an indicator gene at a  
CC rash site and non-rash site of a person with atopic dermatitis or  
CC psoriasis, comparing these levels with those of a healthy person, and  
CC determining that if the levels of indicators are higher or lower, then  
CC this indicates the disease. Also included are a reagent for detecting  
CC atopic dermatitis or psoriasis, a kit for screening for treatments, a  
CC transgenic non human vertebrate animal models for the diseases, an agent  
CC for inducing the diseases in mice and a DNA chip for assaying for the  
CC indicator genes. The method is used for treatment, detection and animal  
CC models for research of atopic dermatitis and psoriasis. The present  
CC sequence is a protein encoded by an indicator gene of the invention.  
XX  
XX Sequence 543 AA;  
SQ  
Query Match 100.0%; Score 2841; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 3.2e-275;  
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQADVDLDFFTQBPPLHLVSPSFLSVT 60  
Db 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQADVDLDFFTQBPPLHLVSPSFLSVT 60  
Qy 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGTKTDFLI FDPKKESTFEERSYQWS 120  
Db 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGTKTDFLI FDPKKESTFEERSYQWS 120  
Qy 121 QVNODICKYGSIPDPVEEKLRLWEPYQEQLLREHYQKKFNKSTYSRSSVDVLYTFANCS 180  
Db 121 QVNODICKYGSIPDPVEEKLRLWEPYQEQLLREHYQKKFNKSTYSRSSVDVLYTFANCS 180  
Qy 181 GLDLIFGNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNPNPSFLKADIFINGS 240  
Db 181 GLDLIFGNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNPNPSFLKADIFINGS 240  
Qy 241 OLGEDFIQLHLKLRKSTFKNAKLYGPDVGOPRRKTAQMLKSLFKAGGEVDSVTWHHYL 300  
Db 241 OLGEDFIQLHLKLRKSTFKNAKLYGPDVGOPRRKTAQMLKSLFKAGGEVDSVTWHHYL 300  
Qy 301 NGRTATREDFLNPDLIDIFTSSVQKVFQVVESTRPGKKWLGETSSAYGGAPLLSDTFA 360  
Db 301 NGRTATREDFLNPDLIDIFTSSVQKVFQVVESTRPGKKWLGETSSAYGGAPLLSDTFA 360  
Qy 361 AGFMWLDKLGSLARMGIEVVMRQVFFGAGNYHLVDENFDLPDYWLSLLFKLVGTKVL 420  
Db 361 AGFMWLDKLGSLARMGIEVVMRQVFFGAGNYHLVDENFDLPDYWLSLLFKLVGTKVL 420  
Qy 421 ASVQSKRRKRLRYLHCTNTDNPYREGDITLVAINLHNTKYLRLPYPFSNQVDKYL 480  
Db 421 ASVQSKRRKRLRYLHCTNTDNPYREGDITLVAINLHNTKYLRLPYPFSNQVDKYL 480  
Qy 481 RPLGPHGLLSKSVQNLGLTLKWVDDQTLPLMEKPLRPGSSLGIPAFYSFFVIRNAKVA 540  
Db 481 RPLGPHGLLSKSVQNLGLTLKWVDDQTLPLMEKPLRPGSSLGIPAFYSFFVIRNAKVA 540  
Qy 541 ACI 543  
Db 541 ACI 543  
RESULT 5  
ID ADM48759 standard; protein; 543 AA.  
XX ADM48759;  
XX  
XX 03-JUN-2004 (first entry)  
XX Human hpa protein #2.  
DE

XX Transgenic animal; heparanase; cancer; viral infection; restenosis;  
KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;  
KW human.  
XX Homo sapiens.  
XX  
XX US2003217375-A1.  
XX 20-NOV-2003.  
XX 24-FEB-2003; 2003US-00371218.  
XX 31-AUG-1998; 98WO-US017954.  
PR 01-MAR-1999; 99US-00258892.  
PR 06-FEB-2001; 2001US-0076874.  
PR 19-NOV-2001; 2001US-00988113.  
XX (ZCHA/) ZCHARIA E.  
PA (VLOD/) VLODAVSKY I.  
PA (METZ/) METZGER S.  
PA (PECK/) PECKER I.  
PA (ILAN/) ILAN N.  
PA (CHAJ/) CHAJEK-SHAUL T.  
PA (GOLD/) GOLDSHMIDT O.  
XX Zcharia E, Vlodavsky I, Metzger S, Pecker I, Ilan N;  
PI Chajek-Shaul T, Goldshmidt O;  
PI WPI; 2004-021918/02.  
DR N-PSDB; ADM48748.  
XX  
XX New transgenic non-human animal expressing heparinase, useful as models  
PT for human disease, such as cancers, viral infection, neurodegenerative  
PT diseases, restenosis, atherosclerosis and pulmonary disorders.  
XX  
XX Example 10; Fig 16; 106pp; English.  
XX  
XX The present invention relates to a transgenic non-human animal whose  
CC genome comprises an exogenous polynucleotide sequence, including a  
CC promoter active in tissues of the non-human, a region encoding a human  
CC heparanase, where the promoter and the region encoding human heparanase  
CC are operably linked in the exogenous polynucleotide such that human  
CC heparanase is expressed in at least a portion of the cells of the non-  
CC human animal. The methods and compositions of the present invention are  
CC useful for the production of transgenic animals expressing heparanase, to  
CC be used as models for human diseases such as cancers, viral infection,  
CC restenosis, neurodegenerative diseases, atherosclerosis and pulmonary  
CC disorders. The present sequence is human hpa protein used in the  
CC exemplification of the invention.  
XX  
SQ Sequence 543 AA;  
Query Match 100.0%; Score 2841; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 3.2e-275;  
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQADVDLDFFTQBPPLHLVSPSFLSVT 60  
Db 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQADVDLDFFTQBPPLHLVSPSFLSVT 60  
Qy 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGTKTDFLI FDPKKESTFEERSYQWS 120  
Db 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGTKTDFLI FDPKKESTFEERSYQWS 120  
Qy 121 QVNODICKYGSIPDPVEEKLRLWEPYQEQLLREHYQKKFNKSTYSRSSVDVLYTFANCS 180  
Db 121 QVNODICKYGSIPDPVEEKLRLWEPYQEQLLREHYQKKFNKSTYSRSSVDVLYTFANCS 180  
Qy 181 GLDLIFGNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNPNPSFLKADIFINGS 240  
Db 181 GLDLIFGNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNPNPSFLKADIFINGS 240



QY 241 QLGEDFIQLHKLARKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGGEVIDSVTWHYYL 300  
| | | | |  
Db 241 QLGEDFIQLHKLARKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGGEVIDSVTWHYYL 300  
| | | | |  
QY 301 NGRTATREDFLNPVDLDFISSVQKVFQVVESTPGKKVWLGETSSAYGGGAPLLSDTFA 360  
| | | | |  
Db 301 NGRTATREDFLNPVDLDFISSVQKVFQVVESTPGKKVWLGETSSAYGGGAPLLSDTFA 360  
| | | | |  
QY 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKKLVGTCKVLM 420  
| | | | |  
Db 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKKLVGTCKVLM 420  
| | | | |  
QY 421 ASVQGSRRKRLRVYLHCTNTDNPRYKEGDLTLAYAINLHNTKYLRLPYPSNKKQVDKYLL 480  
| | | | |  
Db 421 ASVQGSRRKRLRVYLHCTNTDNPRYKEGDLTLAYAINLHNTKYLRLPYPSNKKQVDKYLL 480  
| | | | |  
QY 481 RPLGPHGLLSKSVQLNGLTLLKMWDDQTLPLMEKPLRPGSSGLGPAFYSFFVIRNAKVA 540  
| | | | |  
Db 481 RPLGPHGLLSKSVQLNGLTLLKMWDDQTLPLMEKPLRPGSSGLGPAFYSFFVIRNAKVA 540  
| | | | |  
QY 541 ACI 543  
| | |  
Db 541 ACI 543  
| | |  
RESULT 6  
ADN05074  
ID ADN05074 standard; protein; 543 AA.  
AC ADN05074;  
XX  
DT 01-JUL-2004 (first entry)  
XX  
DE Antipsoriatic protein sequence #716.  
XX  
KW antipsoriatic; gene therapy; psoriasis; diagnosis.  
XX  
OS Homo sapiens.  
XX  
PN WO2004028479-A2.  
XX  
PD 08-APR-2004.  
XX  
PF 25-SEP-2003; 2003WO-US030907.  
XX  
PR 25-SEP-2002; 2002US-0414006P.  
XX  
PA (GETH ) GENENTECH INC.  
XX  
PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;  
PI Wu TD;  
XX  
XX WPI; 2004-305105/28.  
DR N-PSDB; ADN05073.  
XX  
XX New PRO nucleic acid or polypeptide, useful for preparing a  
PT pharmaceutical composition for diagnosing or treating psoriasis in a  
PT mammal.  
XX  
XX Claim 9; SEQ ID NO 1468; 3069pp; English.  
XX  
XX The invention relates to novel polynucleotide and polypeptides for  
CC treating psoriasis or a sequence having at least 80% identity to the  
CC above sequences. The nucleic acid is useful for preparing a composition  
CC for diagnosing or treating psoriasis in a mammal. This sequence  
CC corresponds to one of the polypeptides of the invention.  
XX  
XX Sequence 543 AA;  
SQ  
Query Match 100.0%; Score 2841; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 3.2e-275;  
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQADVVLDLDFFTQEPHLHVSFSLSVT 60  
| | | | |  
Db 1 MLRSKPALPPPLMLLLGLPLSPGALPRPAQADVVLDLDFFTQEPHLHVSFSLSVT 60  
| | | | |  
QY 61 IDANLATDPRFLIILGSPKLRTLARGSLPAYLRFQGTKTDFLIDFKKESTFEERSYQWS 120  
| | | | |  
Db 61 IDANLATDPRFLIILGSPKLRTLARGSLPAYLRFQGTKTDFLIDFKKESTFEERSYQWS 120  
| | | | |  
QY 121 QVNQDICKYGSIPPDVEEKLEWYQEQLLREHYQKKFKNSTYSRSSVDVLYTFANCS 180  
| | | | |  
Db 121 QVNQDICKYGSIPPDVEEKLEWYQEQLLREHYQKKFKNSTYSRSSVDVLYTFANCS 180  
| | | | |  
QY 181 GLDLIFGLNALRLTADLOWNSSNAQLLDYCSSKGVNLSWELGNENPFLKKADIFINGS 240  
| | | | |  
Db 181 GLDLIFGLNALRLTADLOWNSSNAQLLDYCSSKGVNLSWELGNENPFLKKADIFINGS 240  
| | | | |  
QY 241 QLGEDFIQLHKLARKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGGEVIDSVTWHYYL 300  
| | | | |  
Db 241 QLGEDFIQLHKLARKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGGEVIDSVTWHYYL 300  
| | | | |  
QY 301 NGRTATREDFLNPVDLDFISSVQKVFQVVESTPGKKVWLGETSSAYGGGAPLLSDTFA 360  
| | | | |  
Db 301 NGRTATREDFLNPVDLDFISSVQKVFQVVESTPGKKVWLGETSSAYGGGAPLLSDTFA 360  
| | | | |  
QY 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKKLVGTCKVLM 420  
| | | | |  
Db 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKKLVGTCKVLM 420  
| | | | |  
QY 421 ASVQGSRRKRLRVYLHCTNTDNPRYKEGDLTLAYAINLHNTKYLRLPYPSNKKQVDKYLL 480  
| | | | |  
Db 421 ASVQGSRRKRLRVYLHCTNTDNPRYKEGDLTLAYAINLHNTKYLRLPYPSNKKQVDKYLL 480  
| | | | |  
QY 481 RPLGPHGLLSKSVQLNGLTLLKMWDDQTLPLMEKPLRPGSSGLGPAFYSFFVIRNAKVA 540  
| | | | |  
Db 481 RPLGPHGLLSKSVQLNGLTLLKMWDDQTLPLMEKPLRPGSSGLGPAFYSFFVIRNAKVA 540  
| | | | |  
QY 541 ACI 543  
| | |  
Db 541 ACI 543  
| | |  
RESULT 7  
ADN04902  
ID ADN04902 standard; protein; 543 AA.  
AC ADN04902;  
XX  
DT 01-JUL-2004 (first entry)  
XX  
DE Antipsoriatic protein sequence #631.  
XX  
KW antipsoriatic; gene therapy; psoriasis; diagnosis.  
XX  
OS Homo sapiens.  
XX  
PN WO2004028479-A2.  
XX  
PD 08-APR-2004.  
XX  
PF 25-SEP-2003; 2003WO-US030907.  
XX  
PR 25-SEP-2002; 2002US-0414006P.  
XX  
PA (GETH ) GENENTECH INC.  
XX  
PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;  
PI Wu TD;  
XX  
XX WPI; 2004-305105/28.  
DR N-PSDB; ADN04901.  
XX  
XX New PRO nucleic acid or polypeptide, useful for preparing a  
PT pharmaceutical composition for diagnosing or treating psoriasis in a  
PT mammal.

```
PT mammal.
XX Claim 9; SEQ ID NO 1296; 3069pp; English.
XX
XX The invention relates to novel polynucleotide and polypeptides for
XX treating psoriasis or a sequence having at least 80% identity to the
XX above sequences. The nucleic acid is useful for preparing a composition
XX for diagnosing or treating psoriasis in a mammal. This sequence
XX corresponds to one of the polypeptides of the invention.
XX
XX Sequence 543 AA;
XX
XX Query Match 100.0%; Score 2841; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 3.2e-275;
XX Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MLLRSKPAIPPPMLLLGLPLSPGALPRPAQAQDVVDLDFFTQEPHLVSPSFLSVT 60
Db 1 MLLRSKPAIPPPMLLLGLPLSPGALPRPAQAQDVVDLDFFTQEPHLVSPSFLSVT 60
QY 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGTKTDFLIDPKKESTFEERSYQWS 120
Db 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGTKTDFLIDPKKESTFEERSYQWS 120
QY 121 QVNQDICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNSTYSRSSVDVLYTFANCS 180
Db 121 QVNQDICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNSTYSRSSVDVLYTFANCS 180
QY 181 GLDLIFGLNALLRTADLQWNSNAQALLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240
Db 181 GLDLIFGLNALLRTADLQWNSNAQALLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240
QY 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSLKAGGEVIDSVTWHHYLL 300
Db 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSLKAGGEVIDSVTWHHYLL 300
QY 301 NGRTATREDFLNPDVLDIFISSQVKVQVVESTRPGKKVWLGETSSAYGGGAPLLSDTFA 360
Db 301 NGRTATREDFLNPDVLDIFISSQVKVQVVESTRPGKKVWLGETSSAYGGGAPLLSDTFA 360
QY 361 AGFWMLDKLGLSARMGIEVVMRQVFFGAGNHYLVDENFDPLPDYWLSSLFKLVGTQVLM 420
Db 361 AGFWMLDKLGLSARMGIEVVMRQVFFGAGNHYLVDENFDPLPDYWLSSLFKLVGTQVLM 420
QY 421 ASVQSGKRRKRLRVYLHCTNTDNPRIYEGDGLTLVAINLHNVTKYLRLPYPSNKOVDKYL 480
Db 421 ASVQSGKRRKRLRVYLHCTNTDNPRIYEGDGLTLVAINLHNVTKYLRLPYPSNKOVDKYL 480
QY 481 RPLGPHGLLSKSVQLNGLTLKMWDDQTLPLPMEKPLRPGSSSLGLPAPFSYFFVIRNAKVA 540
Db 481 RPLGPHGLLSKSVQLNGLTLKMWDDQTLPLPMEKPLRPGSSSLGLPAPFSYFFVIRNAKVA 540
QY 541 ACI 543
Db 541 ACI 543
RESULT 8
ADQ80372
ID ADQ80372 standard; protein; 543 AA.
XX
AC ADQ80372;
XX
XX 21-OCT-2004 (first entry)
XX Hepatanase protein.
XX cytostatic; epidermal growth factor receptor modulator; identification;
XX therapeutic response; cancer; EGFR; biomarker.
XX Homo sapiens.
XX OS
XX PN WO2004063709-A2.
XX
```

```
XX 29-JUL-2004.
XX 08-JAN-2004; 2004WO-US000368.
XX 08-JAN-2003; 2003US-0438735P.
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX Amler LC, Januario T;
XX WPI; 2004-544114/52.
XX N-PSDB; ADQ80253.
XX
XX Identifying a mammal that will respond therapeutically to a method of
XX treating cancer comprises comparing the level of a biomarker in a mammal
XX before and after exposure to an epidermal growth factor receptor (EGFR)
XX modulator.
XX
XX Disclosure; SEQ ID NO 144; 520pp; English.
XX
XX The invention relates to a method of identifying a mammal that will
XX respond therapeutically to a method of treating cancer by administering
XX an epidermal growth factor receptor (EGFR) modulator by comparing the
XX level of a biomarker in a mammal before and after exposure to an EGFR
XX modulator. The method comprises: (a) measuring, in the mammal, the level
XX of at least one biomarker identified in the specification; (b) exposing
XX the mammal to the EGFR modulator; and (c) measuring in the mammal the
XX level of the biomarker, where a difference in the level in step (c)
XX compared to step (a) indicates that the mammal will respond
XX therapeutically to the method of treating cancer. The method and
XX biomarkers are useful for identifying a mammal that will respond
XX therapeutically to a method of treating cancer by administering an
XX epidermal growth factor receptor (EGFR) modulator. This sequence
XX corresponds to one of the biomarkers whose levels of expression is
XX measured in the method of the invention.
XX
XX Sequence 543 AA;
XX
XX Query Match 100.0%; Score 2841; DB 8; Length 543;
XX Best Local Similarity 100.0%; Pred. No. 3.2e-275;
XX Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MLLRSKPAIPPPMLLLGLPLSPGALPRPAQAQDVVDLDFFTQEPHLVSPSFLSVT 60
Db 1 MLLRSKPAIPPPMLLLGLPLSPGALPRPAQAQDVVDLDFFTQEPHLVSPSFLSVT 60
QY 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGTKTDFLIDPKKESTFEERSYQWS 120
Db 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGTKTDFLIDPKKESTFEERSYQWS 120
QY 121 QVNQDICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNSTYSRSSVDVLYTFANCS 180
Db 121 QVNQDICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNSTYSRSSVDVLYTFANCS 180
QY 181 GLDLIFGLNALLRTADLQWNSNAQALLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240
Db 181 GLDLIFGLNALLRTADLQWNSNAQALLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240
QY 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSLKAGGEVIDSVTWHHYLL 300
Db 241 QLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSLKAGGEVIDSVTWHHYLL 300
QY 301 NGRTATREDFLNPDVLDIFISSQVKVQVVESTRPGKKVWLGETSSAYGGGAPLLSDTFA 360
Db 301 NGRTATREDFLNPDVLDIFISSQVKVQVVESTRPGKKVWLGETSSAYGGGAPLLSDTFA 360
QY 361 AGFWMLDKLGLSARMGIEVVMRQVFFGAGNHYLVDENFDPLPDYWLSSLFKLVGTQVLM 420
Db 361 AGFWMLDKLGLSARMGIEVVMRQVFFGAGNHYLVDENFDPLPDYWLSSLFKLVGTQVLM 420
QY 421 ASVQSGKRRKRLRVYLHCTNTDNPRIYEGDGLTLVAINLHNVTKYLRLPYPSNKOVDKYL 480
Db 421 ASVQSGKRRKRLRVYLHCTNTDNPRIYEGDGLTLVAINLHNVTKYLRLPYPSNKOVDKYL 480
```

Db 421 ASVQSGRRKRLVYLHCTNTDNPRYKEGDLTLVAINLHNTKYLRLPYPSNKKQVKYLL 480

Qy 481 RPLGPHGLSKSVQLNGLTLMQVDDQTLPLMEKPLRPGSSLGLPAPSYFFVIRNAKVA 540

Db 481 RPLGPHGLSKSVQLNGLTLMQVDDQTLPLMEKPLRPGSSLGLPAPSYFFVIRNAKVA 540

Qy 541 ACI 543

Db 541 ACI 543

RESULT 9

ADR88210

ID ADR88210 standard; protein; 543 AA.

AC ADR88210;

XX 18-NOV-2004 (first entry)

DT Human preproheparanase.

DE

XX Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;

KW autoimmune disorder; cancer; angiogenesis; metastatic disease;

KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;

KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;

KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;

KW human; heparanase; enzyme.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Peptide 1. .35

FT /label= Signal\_peptide

FT Protein 36. .543

FT /label= Mature\_heparanase

FT Region 36. .109

FT /note= "8 KDa subunit of mature heparanase dimer"

FT Domain 89. .107

FT /note= "Functional peptide epitope"

FT Region 158. .543

FT /note= "45 KDa subunit of mature heparanase dimer"

FT Domain 219. .233

FT /note= "Functional peptide epitope"

FT Active-site 225

FT /note= "Active site residue"

FT Binding-site 258. .266

FT /note= "Putative heparin binding domain"

FT Domain 294. .307

FT /note= "Functional peptide epitope"

FT Domain 334. .348

FT /note= "Functional peptide epitope"

FT Active-site 343

FT /note= "Active site residue"

FT Binding-site 414. .420

FT /note= "Putative heparin binding domain"

FT Domain 437. .446

FT /note= "Functional peptide epitope"

XX

US2004170631-A1.

XX

02-SEP-2004.

XX

28-NOV-2003; 2003US-00722502.

XX

02-SEP-1997; 97US-00922170.

PR 01-MAY-1998; 98US-00071739.

PR 04-NOV-1998; 98US-00186200.

PR 19-FEB-2003; 2003US-00368044.

PR 22-AUG-2003; 2003US-00645659.

XX

{YACO/} YACOBY-ZEEVI O.

PA (PERE/) PERETZ T.

PA (MIRO/) MIRON D.

PA (SHLO/) SHLOMI Y.

PA (PECK/) PECKER I.

PA (AYAL/) AYAL-HERSHKOVITZ M.

PA (FEIN/) FEINSTEIN E.

PA (VGEL/) VAN GELDER J M.

PA (VLOD/) VLODAVSKY I.

PA (FRIE/) FRIEDMANN Y.

XX

PI Jacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Becker I;

PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;

PI Friedmann Y;

XX

DR WPI; 2004-625084/60.

XX

FT Targeted drug delivery to a heparanase-expressing tissue of a patient,

FT useful for treating heparanase-associated conditions such as inflammation

FT or cancer, comprises administering a drug and an anti-heparanase antibody

FT complex.

FT

PS Claim 2; SEQ ID NO 4; 58pp; English.

XX

CC The invention relates to a method of targeted drug delivery to a tissue

CC of a patient, the tissue expressing heparanase. The method comprises

CC providing a complex of a drug directly or indirectly linked to an anti-

CC heparanase antibody, and administering the complex to the patient. In the

CC targeted drug delivery, the antibody comprises an antibody or its portion

CC capable of specifically binding to at least one epitope of a heparanase

CC protein. The composition and methods of the invention are useful for

CC diagnosing, preventing or treating conditions associated with heparanase

CC catalytic activity (e.g. an inflammatory disorder, wound, scar,

CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell

CC proliferation, invasion of circulating tumour cells and metastatic

CC disease), for purifying heparanase, or for developing drugs for those

CC heparanase-associated conditions. The vasculopathy is atherosclerosis,

CC restenosis or aneurysm. The cancerous condition is a solid cancer or a

CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy

CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous

CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous

CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,

CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and

CC multiple myeloma. The solid cancer is selected from tumours in lip and

CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,

CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,

CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of

CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue

CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,

CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic

CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary

CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,

CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,

CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,

CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's

CC sarcoma. The present sequence is human preproheparanase.

XX

SQ Sequence 543 AA;

Query Match 100.0%; Score 2841; DB 8; Length 543;

Best Local Similarity 100.0%; Pred. No. 3.2e-275;

Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLLRSKPALPPPLMLLLGLPLSPGALPRPAQAQDVVDLDFFTQEPHLVSPFLSVT 60

Db 1 MLLRSKPALPPPLMLLLGLPLSPGALPRPAQAQDVVDLDFFTQEPHLVSPFLSVT 60

Qy 61 IDANLATDPRFLILGLSPKLTARGLSPAYLRFGGTTKDFLIDPKKESTPEERSYMQS 120

Db 61 IDANLATDPRFLILGLSPKLTARGLSPAYLRFGGTTKDFLIDPKKESTPEERSYMQS 120

Qy 121 QVNQDICKYGSIPDVEEKLEWYQEQOLLREHYQKFKNSTYSRSSVDVLYTFANCS 180

Db 121 QVNQDICKYGSIPDVEEKLEWYQEQOLLREHYQKFKNSTYSRSSVDVLYTFANCS 180

Qy 181 GLDLIFGLNALLRTADLQWNSSNAQLLDYCSSKGYNISWELGNPNFSLKKADIFINGS 240

Db 181 GLDLIFGLNALLRTADLQWSSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240  
QY 241 QLGEDFIQLHKLKSTFKNAKLYGPDVGPRKRTAKMLKSFLLKAGGEVDSVTHHYYL 300  
Db 241 QLGEDFIQLHKLKSTFKNAKLYGPDVGPRKRTAKMLKSFLLKAGGEVDSVTHHYYL 300  
QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKWLGESTSSAYGGAPLLSDTFA 360  
Db 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKWLGESTSSAYGGAPLLSDTFA 360  
QY 361 AGFWMLDKGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKLVGTVKVL 420  
Db 361 AGFWMLDKGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKLVGTVKVL 420  
QY 421 ASVQGSRRKRLRVYLHCTNTDNPYKEGDLYALNHNVTYKLRPLYPFSNKQVDKYL 480  
Db 421 ASVQGSRRKRLRVYLHCTNTDNPYKEGDLYALNHNVTYKLRPLYPFSNKQVDKYL 480  
QY 481 RPLGPHGLLSKSVQLNGLTLKWVDDQTLPLMEKPLRPGSSGLGPAFYSFFVIRNAKVA 540  
Db 481 RPLGPHGLLSKSVQLNGLTLKWVDDQTLPLMEKPLRPGSSGLGPAFYSFFVIRNAKVA 540  
QY 541 ACI 543  
Db 541 ACI 543

RESULT 10  
ADP25079  
ID ADP25079 standard; protein; 543 AA.  
XX  
AC ADP25079;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE PRO polypeptide SEQ ID NO:2257.  
XX  
KW PRO; antiinflammatory; antiarthritic; antirheumatic; immunosuppressive;  
KW osteopathic; antidiabetic; dermatological; antipsoriatic; antiallergic;  
KW antiasthmatic; hepatotropic; respiratory; gene therapy; immune system.  
OS Unidentified.  
XX  
PN WO2004041170-A2.  
XX  
PD 21-MAY-2004.  
XX  
PF 30-OCT-2003; 2003WO-US034312.  
XX  
PR 01-NOV-2002; 2002US-0423394P.  
XX  
XX (GETH ) GENENTECH INC.  
XX  
PA Clark H, Schoenfeld J, Van Lookeren M, Williams PM, Wood WT;  
PI Wu TD;  
PI WPI; 2004-419628/39.  
DR N-PSDB; ADP25078.  
XX  
XX New PRO polypeptides and polynucleotides, useful for treating e.g.  
PT erythematous, rheumatoid arthritis, diabetes mellitus, immune-mediated  
PT renal disease, or demyelinating diseases of the central or peripheral  
PT nervous system.  
XX  
PS Claim 7; SEQ ID NO 2257; 2940pp; English.  
XX  
XX The invention relates to a novel isolated nucleic acid and the PRO  
CC polypeptide encoded by it. A protein of the invention has  
CC antiinflammatory, antiarthritic, antirheumatic, immunosuppressive,  
CC osteopathic, antidiabetic, dermatological, antipsoriatic, antiallergic,  
CC antiasthmatic, hepatotropic, and respiratory activity. A polynucleotide  
CC of the invention may have a use in gene therapy. The PRO polypeptide, its

CC agonist, antagonist, or antibody that specifically binds to the  
CC polypeptide is useful for treating an immune related disorder such as  
CC systemic lupus erythematosus, rheumatoid arthritis, osteoarthritis,  
CC juvenile chronic arthritis, a spondyloarthropathy, systemic sclerosis, an  
CC idiopathic inflammatory myopathy, Sjogren's syndrome, systemic  
CC vasculitis, sarcoidosis, autoimmune haemolytic anaemia, autoimmune  
CC thrombocytopenia, thyroiditis, diabetes mellitus, immune-mediated renal  
CC disease, a demyelinating disease of the central or peripheral nervous  
CC system, idiopathic demyelinating polyneuropathy, Guillain-Barre syndrome,  
CC a chronic inflammatory demyelinating polyneuropathy, a hepatobiliary  
CC disease, infectious or autoimmune chronic active hepatitis, primary  
CC biliary cirrhosis, granulomatous hepatitis, sclerosing cholangitis,  
CC inflammatory bowel disease, gluten-sensitive enteropathy, Whipple's  
CC disease, an autoimmune or immune-mediated skin disease, a bullous skin  
CC disease, erythema multiforme, contact dermatitis, psoriasis, an allergic  
CC disease, asthma, allergic rhinitis, atopic dermatitis, food  
CC hypersensitivity, urticaria, an immunologic disease of the lung,  
CC eosinophilic pneumonia, idiopathic pulmonary fibrosis, hypersensitivity  
CC pneumonitis, a transplantation associated disease, graft rejection or  
CC graft-versus-host disease. The present sequence represents a PRO protein  
CC of the invention.  
XX  
SQ Sequence 543 AA;

Query Match 100.0%; Score 2841; DB 8; Length 543;  
Best Local Similarity 100.0%; Pred. No. 3.2e-275;  
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRSKDALPPPLMLLLGLPLSPGALPRPAQADVDLDFETORPLHLVSPFLSVT 60  
Db 1 MLRSKDALPPPLMLLLGLPLSPGALPRPAQADVDLDFETORPLHLVSPFLSVT 60  
QY 61 IDANLATDPRFLILGSPKLTARLSPAYLRFGTKTDFLIIDPKKESTFEERSYQWS 120  
Db 61 IDANLATDPRFLILGSPKLTARLSPAYLRFGTKTDFLIIDPKKESTFEERSYQWS 120  
QY 121 QVNQDICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNSTYSSRSSVDVLTFFANCS 180  
Db 121 QVNQDICKYGSIPDPVEEKLRLWPYQEQLLREHYQKFKNSTYSSRSSVDVLTFFANCS 180  
QY 181 GLDLIFGLNALLRTADLQWSSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240  
Db 181 GLDLIFGLNALLRTADLQWSSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240  
QY 241 QLGEDFIQLHKLKSTFKNAKLYGPDVGPRKRTAKMLKSFLLKAGGEVDSVTHHYYL 300  
Db 241 QLGEDFIQLHKLKSTFKNAKLYGPDVGPRKRTAKMLKSFLLKAGGEVDSVTHHYYL 300  
QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKWLGESTSSAYGGAPLLSDTFA 360  
Db 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKWLGESTSSAYGGAPLLSDTFA 360  
QY 361 AGFWMLDKGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKLVGTVKVL 420  
Db 361 AGFWMLDKGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKLVGTVKVL 420  
QY 421 ASVQGSRRKRLRVYLHCTNTDNPYKEGDLYALNHNVTYKLRPLYPFSNKQVDKYL 480  
Db 421 ASVQGSRRKRLRVYLHCTNTDNPYKEGDLYALNHNVTYKLRPLYPFSNKQVDKYL 480  
QY 481 RPLGPHGLLSKSVQLNGLTLKWVDDQTLPLMEKPLRPGSSGLGPAFYSFFVIRNAKVA 540  
Db 481 RPLGPHGLLSKSVQLNGLTLKWVDDQTLPLMEKPLRPGSSGLGPAFYSFFVIRNAKVA 540  
QY 541 ACI 543  
Db 541 ACI 543

RESULT 11  
ADT78177  
ID ADT78177 standard; protein; 543 AA.  
XX

AC ADT78177;  
 XX  
 DT 13-JAN-2005 (first entry)  
 XX  
 DE Human heparanase protein.  
 XX  
 KW Antibody; epitope; heparanase; pathological condition; angiogenesis;  
 KW cell proliferation; cancerous condition; tumour cell invasion;  
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;  
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;  
 KW cytostatic; antiinflammatory; vulnerable; antiarteriosclerotic;  
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Binding-site 157..162  
 FT Binding-site /note= "Putative heparin binding site"  
 FT Binding-site 271..277  
 FT Binding-site /note= "Putative heparin binding site"  
 FT Binding-site 426..433  
 FT Binding-site /note= "Putative heparin binding site"  
 XX  
 XX US2004213789-A1.  
 PN  
 XX  
 XX 28-OCT-2004.  
 PD  
 XX  
 XX 22-AUG-2003; 2003US-00645659.  
 XX  
 PR 02-SEP-1997; 97US-00922170.  
 PR 01-MAY-1998; 98US-00071739.  
 PR 04-NOV-1998; 98US-00186200.  
 PR 19-FEB-2003; 2003US-00368044.  
 XX  
 XX (YACO/) YACOBY-ZEEVI O.  
 PA (PERE/) PERETZ T.  
 PA (MIRO/) MIRON D.  
 PA (SHLO/) SHLOMI Y.  
 PA (PECK/) PECKER I.  
 PA (AYAL/) AYAL-HERSHKOVITZ M.  
 PA (FEIN/) FEINSTEIN E.  
 PA (GELD/) GELDER J M V.  
 PA (VLOD/) VLODAVSKY I.  
 PA (FRIE/) FRIEDMANN Y.  
 XX  
 XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
 PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;  
 PI Friedmann Y;  
 XX  
 XX WPI; 2004-774790/76.  
 DR  
 XX  
 XX New neutralizing monoclonal anti-heparanase antibodies, useful for  
 PT detecting, treating or preventing cancer, inflammatory or autoimmune  
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.  
 XX  
 XX Claim 5; SEQ ID NO 4; 68pp; English.  
 PS  
 XX  
 XX The invention relates to an isolated antibody or antibody portion capable  
 CC of specifically binding to or elicited by at least one epitope of a  
 CC heparanase protein, where the heparanase protein is at least 60%  
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and  
 CC where at least one epitope comprises a sequence at least 70% homologous  
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)  
 CC a hybridoma cell line comprising a cell line for producing the monoclonal  
 CC antibody, (b) a method for detecting, treating or preventing a  
 CC pathological condition or a heparanase-related disorder or condition in a  
 CC subject, (c) a method for monitoring the state of a heparanase-related  
 CC disorder or condition in a subject, and (d) a pharmaceutical composition  
 CC comprising the isolated anti-heparanase antibody or antibody portion and  
 CC a pharmaceutical carrier. The antibody, methods, and composition are  
 CC useful for detecting, treating, preventing or monitoring a pathological  
 CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition  
 CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,

CC or prostate cancer), minor cell proliferation, invasion of circulating  
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder  
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy  
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or  
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,  
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell  
 CC carcinoma) in a mammal. This sequence represents human heparanase.  
 XX  
 SQ Sequence 543 AA;  
 Query Match 100.0%; Score 2841; DB 8; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-275;  
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MLLRSKPALPPPLMLLLGPIGLSPGALPPRAQAQDVVDLDFFTQEPHLHVPSPFLSVT 60  
 DB 1 MLLRSKPALPPPLMLLLGPIGLSPGALPPRAQAQDVVDLDFFTQEPHLHVPSPFLSVT 60  
 QY 61 IDANLATDPRFLILGSPKRLTARGLSPAYLRFGGTKDTFLIFDPKKESTFEERSYWQS 120  
 DB 61 IDANLATDPRFLILGSPKRLTARGLSPAYLRFGGTKDTFLIFDPKKESTFEERSYWQS 120  
 QY 121 QVNQDICKYGISPPDVEEKLRLWPYQQLLREHYQKFKNSTYSRSSVDVLYTFANCS 180  
 DB 121 QVNQDICKYGISPPDVEEKLRLWPYQQLLREHYQKFKNSTYSRSSVDVLYTFANCS 180  
 QY 181 GLDLIFGLNALRLTADLQWSSNAQLLDYCSSKGYNISWELGNPNPFLKKADIFINGS 240  
 DB 181 GLDLIFGLNALRLTADLQWSSNAQLLDYCSSKGYNISWELGNPNPFLKKADIFINGS 240  
 QY 241 QLGEDFTQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSLKAGGEVIDSVTWHYYL 300  
 DB 241 QLGEDFTQLHKLRLKSTFKNAKLYGPDVGQPRRTAKMLKSLKAGGEVIDSVTWHYYL 300  
 QY 301 NGRTATREDPLNDVLDIFISSVQKVFQVVESTRPGKVKWLGETSSAYGGGAPLLSDTEA 360  
 DB 301 NGRTATREDPLNDVLDIFISSVQKVFQVVESTRPGKVKWLGETSSAYGGGAPLLSDTEA 360  
 QY 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLPDYWLSSLFKLVGTVKVL 420  
 DB 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPLPDYWLSSLFKLVGTVKVL 420  
 QY 421 ASVQGSKKRKLRYVYLHCTNTDNPRYKEGDLTLYAINLHNVTYKLYLPPFPFNQVDKYL 480  
 DB 421 ASVQGSKKRKLRYVYLHCTNTDNPRYKEGDLTLYAINLHNVTYKLYLPPFPFNQVDKYL 480  
 QY 481 RPLGPHGLLSKSVQNLGTLTKWVDDOTLPLMEKPLRPGSSILGPAFYSFFVIRNAKVA 540  
 DB 481 RPLGPHGLLSKSVQNLGTLTKWVDDOTLPLMEKPLRPGSSILGPAFYSFFVIRNAKVA 540  
 QY 541 ACI 543  
 DB 541 ACI 543  
 RESULT 12  
 ADY27036  
 ID ADY27036 standard; protein; 543 AA.  
 XX  
 AC ADY27036;  
 XX  
 DT 05-MAY-2005 (first entry)  
 XX  
 DE Human heparanase protein.  
 XX  
 KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
 KW neurological disease; viral infection; infection; cytostatic;  
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
 KW protease; enzyme; enzyme purification.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2005016227-A2.

XX PD 24-FEB-2005.

XX PF 12-AUG-2004; 2004WO-1L000744.

XX PR 14-AUG-2003; 2003US-0494800P.

XX PR 12-JAN-2004; 2004US-0535492P.

XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX PI Van-Gelder JM, Miron D;

XX DR WPI; 2005-182203/19.

XX Regulating heparanase activity, useful for treating heparanase-associated diseases (e.g. cancer, inflammation, cardiovascular diseases, neurological diseases or viral diseases) comprises modulating heparanase activation.

XX PS Disclosure; SEQ ID NO 8; 211pp; English.

XX CC The invention relates to a method of regulating heparanase activity in a tissue or regulating a biological process depending at least in part on heparanase activity comprising modulating heparanase activation. The invention also relates to methods of treating a heparanase- or heparin binding protein-associated disease or disorder in a subject, a pharmaceutical composition for use in the treatment of a heparanase-associated disease or disorder comprising a therapeutic amount of an agent capable of modulating heparanase activation and a pharmaceutical carrier or diluent, a method of identifying a protease activator of heparanase, a protease substrate mimetic comprising a peptide representing a subset or all substrate residues or cleavage sites of human heparanase or an equivalent non-human heparanase, a method of producing active heparanase and a method of modulating an adhesion activity of heparanase. The composition and methods are useful for modulating heparanase activation and for treating heparanase-associated diseases or disorders such as cancer, inflammation, cardiovascular diseases, neurological diseases or viral infections. This sequence represents a human heparanase protein used in the scope of the invention.

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 2841; DB 9; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-275;  
 Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLRSKPPALPPMLLLGLPLGLSPGALPRPAQAQDVVDLDFPTQPLHLVSPFLSVT 60  
 DB 1 MLLRSKPPALPPMLLLGLPLGLSPGALPRPAQAQDVVDLDFPTQPLHLVSPFLSVT 60

QY 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGKTDFLIDPKKESTFEERSYQWS 120  
 DB 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGGKTDFLIDPKKESTFEERSYQWS 120

QY 121 QVNQDICKYGSIPDVVEKRLWPYQEQLLREHYQKFKFNSTYSRSSVDVLYTFANCS 180  
 DB 121 QVNQDICKYGSIPDVVEKRLWPYQEQLLREHYQKFKFNSTYSRSSVDVLYTFANCS 180

QY 181 GLDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYNISWELGNPNFSFKKADIFINGS 240  
 DB 181 GLDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYNISWELGNPNFSFKKADIFINGS 240

QY 241 QLGEDFIQLHLKLRKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGEVIDSVTHHYLL 300  
 DB 241 QLGEDFIQLHLKLRKSTFKNAKLYGPDVGQPRRTAKMLKSFLLKAGEVIDSVTHHYLL 300

QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKVWLGETSSAYGGCAPLLSDTFA 360  
 DB 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKVWLGETSSAYGGCAPLLSDTFA 360

QY 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPDVLWLSLLPKKLVTGKVL 420  
 DB 361 AGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDENFDPDVLWLSLLPKKLVTGKVL 420

QY 421 ASVQSGRRKRLRVYLHCTNTDNPYKGGDLTLYAINLHNVTKYLRLPYFPFSNKQVDKYL 480  
 DB 421 ASVQSGRRKRLRVYLHCTNTDNPYKGGDLTLYAINLHNVTKYLRLPYFPFSNKQVDKYL 480

QY 481 RPLGPHGLLSKSVOLNGITLQWVDDQTLPLMEKPLRPGSSIGLPAFYSFPVIRNAKVA 540  
 DB 481 RPLGPHGLLSKSVOLNGITLQWVDDQTLPLMEKPLRPGSSIGLPAFYSFPVIRNAKVA 540

QY 541 ACI 543  
 DB 541 ACI 543

RESULT 13  
 AEA42426  
 ID AEA42426 standard; protein; 543 AA.  
 XX  
 AC AEA42426;  
 XX  
 DT 28-JUL-2005 (first entry)  
 XX  
 DE Human heparanase protein SEQ ID NO:4.  
 XX  
 KW antibody; heparanase; antiinflammatory; vulnerrary; immunosuppressive;  
 KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;  
 KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;  
 KW angiogenesis disorder; cancer; tumor; metastasis.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 89..107  
 FT Peptide /note= "heparanase epitope SEQ ID NO:7"  
 FT Peptide 219..233  
 FT Peptide /note= "heparanase epitope SEQ ID NO:8"  
 FT Misc-difference 246  
 FT Peptide /note= "encoded by TAT"  
 FT Peptide 294..307  
 FT Peptide /note= "heparanase epitope SEQ ID NO:10"  
 FT Peptide 334..348  
 FT Peptide /note  
 FT Peptide 437..446  
 FT Peptide /note= "heparanase epitope SEQ ID NO:6"  
 XX  
 PN AUZ004201462-A1.  
 XX  
 PD 06-MAY-2004.  
 XX  
 PF 08-APR-2004; 2004AU-00201462.  
 XX  
 PR 08-APR-2004; 2004AU-00201462.  
 XX  
 PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.  
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.  
 XX  
 PI Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;  
 PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-HersHKovitz M, Ben-Artzi H;  
 PI Feinstein E;  
 XX  
 DR WPI; 2005-173343/19.  
 DR N-PSDB; AEA42434, AEA42435, AEA42460.  
 XX  
 PT Novel isolated antibody capable of specifically binding to epitope of heparanase protein, useful for preventing and treating heparanase-related disorder such as inflammatory disorder, scars, autoimmune conditions or angiogenesis.  
 PT  
 PS Claim 2; SEQ ID NO 4; 260pp; English.  
 XX  
 CC The invention relates to an isolated antibody or its portion (I) capable of specifically binding to an epitope of a heparanase protein. Also described: (i) a cell line (ii) for producing a monoclonal antibody or

its portion, comprising a cell line for producing (1); (2) a pharmaceutical composition comprising (1) and a carrier; and (3) an affinity medium (III) for binding human heparanase polypeptides, comprising (1) immobilized to a chemically inert, insoluble carrier. (1) useful for treating a subject suffering from a pathological condition, which involves administering (1) to the subject. (1) is useful for preventing and treating heparanase-related disorder or condition chosen from inflammatory disorder, wound, scar, vasculopathy, autoimmune condition, angiogenesis, cell proliferation, cancerous condition, tumor cell proliferation, invasion of circulating tumor cells and metastatic disease. (1) is useful for detecting the presence of heparanase polypeptide in a sample. (1) is useful for detecting heparanase-related disease or condition in a subject such as vertebrate, preferably mammal e.g., human. The heparanase-related disorder or condition further includes renal disease or disorder chosen from diabetic nephropathy, glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome and renal cell carcinoma. The present sequence represents human heparanase, which is used in the exemplification of the present invention.

XX SQ Sequence 543 AA;

Query Match 100.0%; Score 2841; DB 9; Length 543;  
Best Local Similarity 100.0%; Pred. No. 3.2e-275;  
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRSKPALPPMLLLGLPLSPGALPRPAQADVDLDFPTQPLHLVSPSFLSVT 60  
DB 1 MLRSKPALPPMLLLGLPLSPGALPRPAQADVDLDFPTQPLHLVSPSFLSVT 60  
QY 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGCTKTDFLIPDPKKESTFEERSYQWS 120  
DB 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGCTKTDFLIPDPKKESTFEERSYQWS 120  
QY 121 QVNQDICKYGISPPDVEEKLREWPYQEQLLLRHYQKKFNSTYSRSSVDVLYTFANCS 180  
DB 121 QVNQDICKYGISPPDVEEKLREWPYQEQLLLRHYQKKFNSTYSRSSVDVLYTFANCS 180  
QY 181 GLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240  
DB 181 GLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240  
QY 241 QLGEDFIQLHLKLRKSTFKNAKYGPDVGQPRRKTAQMLKSFLLKAGGEVDSVTWHYYL 300  
DB 241 QLGEDFIQLHLKLRKSTFKNAKYGPDVGQPRRKTAQMLKSFLLKAGGEVDSVTWHYYL 300  
QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPKKWLGETSAYGCGAPLLSDTEA 360  
DB 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPKKWLGETSAYGCGAPLLSDTEA 360  
QY 361 AGFWMLDKLGLSARMGIEVVMROVFFGAGNHYLVDENFDPLPDYWL SLLFKKLVTGKVL 420  
DB 361 AGFWMLDKLGLSARMGIEVVMROVFFGAGNHYLVDENFDPLPDYWL SLLFKKLVTGKVL 420  
QY 421 ASYQGSRRKRLRYLHCTNTDNPRYKEDLTLYAINLHNVTYKLRPLYPFSNKQVDKYL 480  
DB 421 ASYQGSRRKRLRYLHCTNTDNPRYKEDLTLYAINLHNVTYKLRPLYPFSNKQVDKYL 480  
QY 481 RPLGPHGLLSKSVQLNGLTLKMWDDQTLPLPMEKPLRPGSSGLPAPSYSEFFVIRNAKVA 540  
DB 481 RPLGPHGLLSKSVQLNGLTLKMWDDQTLPLPMEKPLRPGSSGLPAPSYSEFFVIRNAKVA 540  
QY 541 ACI 543  
DB 541 ACI 543

RESULT 14

AAY30124

ID AAY30124 standard; protein; 588 AA.

XX

AC AAY30124;

XX

DT 20-MAR-2003 (revised)  
DT 14-OCT-1999 (first entry)  
XX  
DE A human protein with heparanase activity.  
XX  
KW Human; heparanase; heparan sulfate; trauma; autoimmune disease;  
KW skin disease; cardiovascular disease; nervous system disease;  
KW Alzheimer's disease; cancer; cancer metastasis; angiogenesis;  
KW inflammation; arthritis.  
XX  
OS Homo sapiens.  
XX WO9940207-A1.  
PN  
PD 12-AUG-1999.  
XX  
PF 05-FEB-1999; 99WO-EP000777.  
XX  
PR 09-FEB-1998; 98GB-00002725.  
XX  
XX (NOVS ) NOVARTIS AG.  
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
PI Nakajima M, Toyoshima M;  
XX WPI; 1999-494300/41.  
DR N-PSDB; AAX86671.  
XX  
PT New heparanase polypeptide useful for treating autoimmune diseases, skin  
PT diseases, cardiovascular diseases and nervous system diseases including  
XX Alzheimer's disease.  
PS Claim 3; Page 29-31; 40pp; English.  
XX  
CC The present sequence represents a polypeptide with human heparanase  
CC biological activity. Antagonists and inhibitors of the protein prevent it  
CC from degrading the extracellular matrix and releasing heparan sulfate  
CC from the extracellular matrix surface. The heparanase protein or the anti  
CC -heparanase antibody are used in pharmaceutical compositions for treating  
CC warm blooded animals suffering from a disease resulting from shortage or  
CC lack of the heparanase protein, or from excessive activity or over-  
CC expression of the heparanase protein, respectively. The heparanase  
CC protein is used in treating diseases such as trauma, autoimmune disease,  
CC skin diseases, cardiovascular diseases and nervous system diseases  
CC including Alzheimer's disease resulting from shortage or lack of  
CC polypeptide. The anti-heparanase antibody is used in treating the  
CC diseases like cancer, cancer metastasis, angiogenesis and inflammation  
CC including arthritis, resulting from excessive activity or over expression  
CC of heparanase protein. The anti-heparanase antibody can be used to detect  
CC the presence or absence of polypeptide and its concentration. (Updated on  
CC 20-MAR-2003 to correct PA field.)  
XX  
SQ Sequence 588 AA;

Query Match 100.0%; Score 2841; DB 2; Length 588;  
Best Local Similarity 100.0%; Pred. No. 3.7e-275;  
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRSKPALPPMLLLGLPLSPGALPRPAQADVDLDFPTQPLHLVSPSFLSVT 60  
DB 46 MLRSKPALPPMLLLGLPLSPGALPRPAQADVDLDFPTQPLHLVSPSFLSVT 105  
QY 61 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGCTKTDFLIPDPKKESTFEERSYQWS 120  
DB 106 IDANLATDPRFLILGSPKLTLAGLSPAYLRFGCTKTDFLIPDPKKESTFEERSYQWS 165  
QY 121 QVNQDICKYGISPPDVEEKLREWPYQEQLLLRHYQKKFNSTYSRSSVDVLYTFANCS 180  
DB 166 QVNQDICKYGISPPDVEEKLREWPYQEQLLLRHYQKKFNSTYSRSSVDVLYTFANCS 225  
QY 181 GLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240  
DB 226 GLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 285



QY 241 QLGEDFIQLHKLKSTFKNAKLYGPDVGQPRKTKAKMLKSFLLKAGGEVIDSVTHHYLL 300  
 |||||  
 Db 286 QLGEDFIQLHKLKSTFKNAKLYGPDVGQPRKTKAKMLKSFLLKAGGEVIDSVTHHYLL 345  
 |||||  
 QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA 360  
 |||||  
 Db 346 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA 405  
 |||||  
 QY 361 AGFMWLDKGLSARMGIEVVMQVFFGAGNYHLVDENFDPLPDYWLSSLLEKKLVGTVLM 420  
 |||||  
 Db 406 AGFMWLDKGLSARMGIEVVMQVFFGAGNYHLVDENFDPLPDYWLSSLLEKKLVGTVLM 465  
 |||||  
 QY 421 ASVQGSRRKRLRYLHCTNTDNPRYKEGDLTYAINLHNVTYKLRYPYPSNKKQVDKYL 480  
 |||||  
 Db 466 ASVQGSRRKRLRYLHCTNTDNPRYKEGDLTYAINLHNVTYKLRYPYPSNKKQVDKYL 525  
 |||||  
 QY 481 RPLGPHGLLSKSVQLNGLTLKMVDDQTLPLMEKPLRPGSSGLPAPFSYFFVIRNAKVA 540  
 |||||  
 Db 526 RPLGPHGLLSKSVQLNGLTLKMVDDQTLPLMEKPLRPGSSGLPAPFSYFFVIRNAKVA 585  
 |||||  
 QY 541 ACI 543  
 |||||  
 Db 586 ACI 588

## RESULT 15

AA02345

ID AAY02345 standard; protein; 543 AA.

XX

AC AAY02345;

XX

DT 09-JUL-1999 (first entry)

XX

DE A human heparanase protein.

XX

KW Heparanase; hp; modulator; heparin-binding growth factor;

KW cellular response; cytokine; cell interaction; plasma lipoprotein;

KW cellular susceptibility; infection; disintegration;

KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;

KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;

KW plasma heparin; micrometastasis; autoimmune lesion; renal failure.

XX

OS Homo sapiens.

XX

PN WO9911798-A1.

XX

PD 11-MAR-1999.

XX

PF 31-AUG-1998; 98WO-US017954.

XX

XX 02-SEP-1997; 97US-00922170.

PR

PR 02-JUL-1998; 98US-00109386.

XX

XX (INSI-) INSIGHT STRATEGY &amp; MARKETING LTD.

PA (HADA-) HADASIT MEDICAL RES SERVICES &amp; DEV.

PA

XX (FRIE-) FRIEDMAN M M.

XX

XX Pecker I, Vlodavsky I, Feinstein E;

XX

XX WPI: 1999-302255/25.

DR

DR N-PSDB; AAX35648.

XX

XX New human polynucleotide useful for treating angiogenesis, restenosis,

PT and inflammation.

XX

XX Claim 6; Fig 1; 63pp; English.

PS

XX The specification describes a polypeptide having heparanase (hp)

CC activity. The recombinant protein is used as a modulator of heparin-

CC binding growth factors, cellular responses to heparin-binding growth

CC factors and cytokines, cell interaction with plasma lipoproteins,

CC cellular susceptibility to viral, protozoal and bacterial infections or

CC disintegration of neurodegenerative plaques. Heparanase may be useful for  
 CC conditions such as wound healing, angiogenesis, restenosis,  
 CC atherosclerosis, inflammation, neurodegenerative diseases, and viral  
 CC infections. Mammalian heparanase can be used to neutralize plasma  
 CC heparin, and anti-heparanase antibodies may be applied for  
 CC immunodetection and diagnosis of micrometastases, autoimmune lesions, and  
 CC renal failure in biopsy specimens, plasma samples, and body fluids. The  
 CC present sequence represents human heparanase

XX

SQ Sequence 543 AA;

Query Match 99.9%; Score 2838; DB 2; Length 543;

Best Local Similarity 99.8%; Pred. No. 6.4e-275;

Matches 542; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLARSKPALPPMLLLGLPLSPGALPRPAQADVDLDFTQRPDLHLVSPSFLSVT 60

Db 1 MLARSKPALPPMLLLGLPLSPGALPRPAQADVDLDFTQRPDLHLVSPSFLSVT 60

QY 61 IDANLATDPRFLILGLSPKLTLAGLSPAYLRFGGTKTDFLIIDPKKKESTFEERSYQWS 120

Db 61 IDANLATDPRFLILGLSPKLTLAGLSPAYLRFGGTKTDFLIIDPKKKESTFEERSYQWS 120

QY 121 QVNODICKYGSIPPDVEEKLRLWPYQEQLLLRHYQKKFNKSTYSRSSVDVLYTFANCS 180

Db 121 QVNODICKYGSIPPDVEEKLRLWPYQEQLLLRHYQKKFNKSTYSRSSVDVLYTFANCS 180

QY 181 GLDLIFGLNALLRTADLOWNSSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240

Db 181 GLDLIFGLNALLRTADLOWNSSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240

QY 241 QLGEDFIQLHKLKSTFKNAKLYGPDVGQPRKTKAKMLKSFLLKAGGEVIDSVTHHYLL 300

Db 241 QLGEDFIQLHKLKSTFKNAKLYGPDVGQPRKTKAKMLKSFLLKAGGEVIDSVTHHYLL 300

QY 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA 360

Db 301 NGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA 360

QY 361 AGFMWLDKGLSARMGIEVVMQVFFGAGNYHLVDENFDPLPDYWLSSLLEKKLVGTVLM 420

Db 361 AGFMWLDKGLSARMGIEVVMQVFFGAGNYHLVDENFDPLPDYWLSSLLEKKLVGTVLM 420

QY 421 ASVQGSRRKRLRYLHCTNTDNPRYKEGDLTYAINLHNVTYKLRYPYPSNKKQVDKYL 480

Db 421 ASVQGSRRKRLRYLHCTNTDNPRYKEGDLTYAINLHNVTYKLRYPYPSNKKQVDKYL 480

QY 481 RPLGPHGLLSKSVQLNGLTLKMVDDQTLPLMEKPLRPGSSGLPAPFSYFFVIRNAKVA 540

Db 481 RPLGPHGLLSKSVQLNGLTLKMVDDQTLPLMEKPLRPGSSGLPAPFSYFFVIRNAKVA 540

QY 541 ACI 543

Db 541 ACI 543

Search completed: June 5, 2006, 12:09:39

Job time : 111.256 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:10:07 ; Search time 18.0785 Seconds  
(without alignments)  
2889.939 Million cell updates/sec

Title: US-10-645-659A-4  
Perfect score: 2841  
Sequence: 1 MLRSKPALPPPLMLLLGP.....LPASYSFFVIRNAKVACI 543  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: piri:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	893.5	31.5	480	JC7506	heparanase protein
2	417	14.7	521	T45608	hypothetical prote
3	169.5	6.0	190	T01953	hypothetical prote
4	112	3.9	2298	T49648	hypothetical prote
5	111	3.9	670	T10666	hypothetical prote
6	111	3.9	688	S32961	hypothetical prote
7	110.5	3.9	356	F64383	hypothetical prote
8	107.5	3.8	575	T12094	beta-fructofuranos
9	106	3.7	670	T38446	microtubule-associ
10	105.5	3.7	411	T57460	hypothetical prote
11	105.5	3.7	879	E91031	probable outer mem
12	104.5	3.7	788	S00652	phosphoribosylamin
13	104	3.7	432	F70411	adenylosuccinate s
14	104	3.7	796	D97065	transketolase [imp
15	104	3.7	2013	A11489	probable peptidogl
16	103.5	3.6	500	D87541	beta-xylosidase [i
17	103.5	3.6	676	AF1153	transcription anti
18	102	3.6	879	F85875	probable fibrinoly
19	102	3.6	897	G02529	dynein heavy chain
20	102	3.6	4644	A38905	dynein heavy chain
21	101.5	3.6	746	T46821	siderophore recept
22	101.5	3.6	746	D82920	RhtA Rhizobactin r
23	101	3.6	594	A82913	hypothetical prote
24	100.5	3.5	604	E75119	hypothetical prote
25	100.5	3.5	687	F85188	retrotransposon li
26	100.5	3.5	847	AG1001	nitrite reductase
27	100.5	3.5	1314	D319488	probable membrane
28	100.5	3.5	1734	D41101	phorbol ester-bind
29	100	3.5	578	B89045	protein B0238.7 [i

RESULT 1  
JC7506  
heparanase protein 2a - human  
C:Species: Homo sapiens (man)  
C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 09-Jul-2004  
C:Accession: JC7506  
R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hirtcock, M.; Pat  
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000  
A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family me  
A:Reference number: JC7506  
A:Accession: JC7506  
A:Molecule type: mRNA  
A:Residues: 1-480 <MCK>  
A:Cross-references: UNIPROT:Q9HB39; UNIPARC:UPI000003E88A; GB:AF282885  
A:Comment: This protein, a intracellular membrane-bound enzyme, has biological and therai  
therapies.  
C:Genetics:  
A:Gene: hpa2a  
A:Map position: 10q23-10q24  
C:Keywords: heparin binding; membrane bound

Query Match	31.5%	Score	893.5	DB 2	Length	480
Best Local Similarity	35.8%	Pred. No.	8.2e-59			
Matches	201	Conservative	75	Mismatches	146	Indels 139; Gaps 9
Qy	20	PLGLPSGAL	-----PRPA-----QAQDVVDLDFEQPLHLVSPS	55		
Db	18	PPACLAFCALYLALLHLHLSLSSQAGDRRLPVDRRAAGLKEKTLILLDVSVKRPVTVNEN	77			
Qy	56	FLSVTIDANLATDPRFLILGLSPKLTLAGLSPAYLRFGGTKTDFLIF----	DPKKEST	111		
Db	78	FLSLQLDPSIIHD-GWLDLFLSSKRLVLTARGLSPAFLRFGGKRTDFLOPQNLRNPAKSR-	135			
Qy	112	FEERSYQSQVNQDICKYISPPDVEEKRLLEWYPQEQLLREHYQKFKPNSTSRSSVD	171			
Db	136	-----GGFGPD-----YYLKNYE-----	148			
Qy	172	VLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYNISWELGNEPNSFLK	231			
Db	149	-----DEPNRYT	156			
Qy	232	KADIFINGSQIGEDFIQLHKLRLK-STFNKAKLYGPDVGQPRKTKAKLKSFLKAGGEVI	290			
Db	157	MHGRAVNSQLGKDYIQLKSLQPIRIYSRASLYGNIGRPRKNVIALLDGFMKVGASTV	216			
Qy	291	DSVTWHYLLNGRTATREDFLNPDVLFISSVQKVFQVVESTREPKKVKWLGETSAYGG	350			
Db	217	DAVTWQHICYIDGRVVKVWMDFLKTRLLDLSQDIRKIQKVNTYTPGKKLWLSGVVTTAG	276			
Qy	351	GAPLLSDTFAAGFMWLDKGLSARMGIEVVRQVFFGAGNYHLVDENFDPLPDYWLISLLF	410			
Db	277	GTNNLSDSYAAGFLWNLTLGMLANQIDVVIHRSFFDHGYNHLVDQNFNPLPDYWLISLLY	336			

Qy	411	KKLVGTVKVLMA	SVQSGSKR-----	KLRYVLHCTNTD	NPYKEGDUTL	VAINLHNT	461
Db	337	KLITGPKVLA	VHVAGLQKPR	PRGPRVIRD	KLRIYAHCTN	HHNHYVRSITL	FTIILHRSR 396
Qy	462	KYLRLPFPFS	KNQVDKYL	RLPLGPHGLLS	KSQVLNGLTL	KQVDDQTL	PLPMKPLRPGSS 521
Db	397	KKIKLAGTL	RDKLVHQY	LLQPVYQ	BGLSKSVQ	LNGQPLVM	VDGTLPELKPRPLRAGT 456
Qy	522	LGLP	AFSFSF	FI	RNAKVAAC	542	
Db	457	LVIP	PVTMG	FFW	KKNV	NALAC	477

## RESULT 2.

A:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 09-Jul-2004  
 C:Accession: T45608  
 R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba  
 submitted to the Protein Sequence Database, December 1999  
 A:Reference number: Z23009  
 A:Accession: T45608  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-521 <BEV>  
 A:Cross-references: UNIPROT:Q9SDA1; UNIPARC:UPI000000A497C; EMBL:AL133421  
 A:Experimental source: cultivar Columbia; BAC clone F13G24  
 C:Genetics:  
 A:Map position: 5  
 A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3  
 A:Note: F13G24.30

Query Match	14.7%	Score 417;	DB 2;	Length 521;
Best Local Similarity	29.2%	Pred Log 3.4e-23;		
Matches 154; Conservative		68; Mismatches 104;	Indels 122;	Gaps 24;
QY	75	LGSPKLTLAGLSPAYLRGCTKTDLPI	EDPKKESTFEERSYQWSQVNDICKVGSI	PP 134
db	55	LTPPLTKAIKAFKPLRIGGSLQDVYDVG	NLKT-----PCR-----	94

Qy	135	DVEKLRLEWPQEQILLREHYKKFNS	---TYSRSSV---	DVLTPANCGLDLIP	186
		:	:		
Db	95	-----PQGM-----	-----NSGLFGSGCLHMKRWDELNSF	---TATGAVVTF	132

Qy	187	GINALLRTADLQ-----WNSNAQLLLDYCSSKGYNI-SWELGNEPNSFLKKADIFIN	238
			:
Db	133	GINALGRHKLKRGKAWGGAWDHINTODFLNYTVSKYVIDSWFEGNEISG--SCVGASVS	190
			:

Qy 239 GSQLGDFIQLHKLLRSTFFKNAKLYGPDVGQP-----RRKTAQMLKSLFKAGEVIDSV 239

DB	191	ASLIGKULILVKDVINK-VIKSNWLRHLLVAEGGF	IEQQWIKLDEL---	SQFSVLDVV	238
Qy	294	TWHYYLNGRT--ATREDLPNPVLDIFISSVKVF	-----QVVSTRGKKVWLGETSSA	347	

Db	THHHYINLGSNDPALVKKIMDPSS---YLSQVSKTFKDVNQTIIQEHGWPASPWVGESGGA	302
	247	
Qy	YGGGAPLLSDTPAAGFMWLDKLGLSARMGIEVVMRQVFSGAGNYHLVDE-NFDPLPDYWL	406
	348	

Db	303	YNSGGRHVSDTFIDSFWYLDQGSARHNTKVYCRQTLVG- GFYGLLEKGTVPNPDIYS	361
Qy	407	SLLFKVLGTVKULMASVOGSKRKLRYVLHCITNDNPRYKEGDLTYAINLHNVTKYL--	464

Db	362	ALLWHRLMGKGVLLAVQTDGPP	-----QLRVYAHCK-----	GRAGVTLLLLNLSNQSDFTVS	413
QY	465	-----	-----RLPYPFs-----	NKQVDKYLLRP-----	495

Db 414 VSGINVLNAESRKKSLDLTKRPFSGWIGSKASDGVLYNREYHLTPENGVLRSKTMVL 473  
::  
Ov 496 NGITIKVVDQDTI.PPI.MEKPI.RP.-GSSTIGLPAFSYSFFVIRNAKVAAC 542

474 NGKSLKPTATGIPSL-EPVLRSVNSPLNVLPLSMFSFVLPNFDASAC 520

### RESULTS

T01953  
hypothetical protein T2L5.6 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999  
C/Accession: T01953  
R/Gaisel, C.; Smith, A.; Le, T.  
submitted to the EMBL Data Library, October 1998  
A/Description: The sequence of A. thaliana T2L5.  
A/Reference number: Z14470  
A/Accession: T01953  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-190 <GBI>  
A/Cross-references: UNIPROT:O82604; UNIPARC:UPI000  
A/Experimental source: cultivar Columbia  
C/Genetics:  
A/Map position: 4  
A/Introns: 36/2; 69/3  
A/Note: T2L5.6  
C/Suprafamily: Arabidopsis thaliana hypothetical p

Query Match	6.0%;	Score	169.5;	DB 2;	Length	190;			
Best Local Similarity	27.8%;	Pred. No.	2.2e-05;						
Matches	54;	Conservative	34;	Mismatches	57;	Indels	49;	Gaps	9

Qy		382	RQVFFGAGNYHLVD - ENFDPLDPDYLSLLFKCLVGTKVLMASVOGSRRKRLRVYLHCTNT	440
Dd		12	ROSLIG-GNYGLLNTFTNPNDYYSALIRQLMGKALFTTSGTK--KIRSTHCA--66	

QY 441 DNPYKEGDLTVAINLHV-----TKYLRLPYPFNSKQVDKYLLRPL 483

Qy 484 GPGLL-----SKSVQLNGLLTKKVVDDDTLPPLMEKPLRPGSSIGLPAF 528  
||::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|  
67 ---KQNS--IIVDEENHIIAIVAVVLENNSTSENIAERK---SARGGDQGLG 111

QY	529	YSFFVIRNAKVAAC	542
		:	
DB	116	GFNGVQREEEYHLTAKDGNLHQSMTLLNGNALQVMSMDGLPPPIPIHINSTEPIIAPYS	177

## RESULT. T 4

A:Accession: T49648  
 A:Reference number: Z25022  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2398 <SCH>  
 A:Cross-references: UNIPROT:Q96U00; UNIPARC:UPI000017B4E4; EMBL:U00001.1  
 A:Experimental source: BAC clone B8B20; strain OK74A  
 C:Genetics:  
 A:Gene: NCSP:B8B20.20  
 A:Map position: 6  
 A:Introns: 426/3

Query Match 3.9%; Score 112; DB 2; Length 2298;  
Best Local Similarity 19.3%; Pred. NO. 19;  
Matches 114; Conservative 79; Mismatches 190; Indels 20

[illegible]

[illegible]



Qy 107 KKESTFEERSYQWQVQNDICKYGSIPDPVEEKRLWEPYOELLRLHRYQKFKNSTYS 166  
Db 73 KRPSVVKSR-----KKGSENIISNMEKTKAIKQKSRRPSKFRSLARPLCITPIDSSPT 128  
Qy 167 RSSVDLYTFANGSGLD-LIFGLNALRTADLOWNSNAQLL-LDYCSSKGYNIISWELGN 224  
Db 129 KTA--TFYTSITTENDELNFSTEELSSFDITLLNSDTSKLSGLDDSSFMEEFVWQVDN 186  
Qy 225 -----EPNSFLKK-----ADIFINGSQIGEDFIQLHKULRK-----ST 257  
Db 187 VLQCEKCKFTPHSGSYLKENLKSRLKGRLDLMLCENTALKEKIDKLNLEKEVPEQLT 246  
Qy 258 FKNAKLYGPDVGOPRR-KTAKMLKSLFKAGGEV-----IDSVTWHHYVLNGRT 304  
Db 247 FLRSK---NSIEKPRNFRKEFLKFLMAQKEIKYLKRKLOIRKIPNYKYSRDLNSKRT 303  
Qy 305 ATREDPLNPVDLD---IFISSQVKVQVVESTRFGKKVWLQETSSAYGGGAPLLSDTFAA 361  
Db 304 PKSQDNWTTQTVPSSLLGVSEVSKVLQL-----KQVQVDITE----- 340  
Qy 362 GFMWLDKLGLSARMGTEVNRQVFFGAGN-----YHLVDENPDPLPDYWLSLFKKLVGTK 417  
Db 341 -LVKIPKPNFSEKLTISNVRNLYNPVGSLDLOFSLTNENF---VHMNSTVYQELLNLK 395  
Qy 418 VLMASSVQGSKRKK 430  
Db 396 SNNSSVDGVKTRR 408

RESULT 10  
S74760  
hypothetical protein slr1617 - *Synechocystis* sp. (strain PCC 6803)  
C/Species: *Synechocystis* sp.  
A/Variety: PCC 6803  
C/Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C/Accession: S74760  
R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, N. A.; Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp.  
A/Reference number: S74322; MUID:97061201; PMID:8905231  
A/Accession: S74760  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-411 <KAN>  
A/Cross-references: UNIPROT:P72895; UNIPARC:UPI000000C0C3B; EMBL:D90901; GB:AB001339; NID  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match	3.7%;	Score 105.5;	DB 2;	Length 411;
Best Local Similarity	21.4%;	Pred. No. 4.3;		
Matches	66;	Conservative 54;	Mismatches 95;	Indels 93; Gaps 17;
Qy	159	KFKNSTYSSRSVDVLYTFANCGLDLIF---	GLNALLRTADLWNSSNAQL-----	206
	:	:	:	:
	:	:	:	:
	:	:	:	:
Db	151	EFRLISPTREQIDI---	FAGSTKLDLLASENDICIVHLANPRVYTSNVAMGQTLTMLRN	207
	:	:	:	:
	:	:	:	:
	:	:	:	:
Qy	207	LLDYCSSKG---	YNISWEL-----GNEPNSFLKK-----	ADIFINGSQLGE 244
	:	:	:	:
	:	:	:	:
	:	:	:	:
Db	208	VIDVCLAKDIPUIYSSWBIYSGYAGTIHADESTPALPRPGYETKYLAETLI		260
	:	:	:	:
	:	:	:	:
	:	:	:	:
Qy	245	DFIQLHLKLRSTFKNAKLYGPDVGOPRRRTAKMLKFLKAGGEVIDSVTWHYYLNGRT		304
	:	:	:	:
	:	:	:	:
	:	:	:	:
Db	261	DHCRRTRGLRCAILRSSPVYGSMDRP-----	KFIFNFFKKASQGQKIVT--HHYING--	311
	:	:	:	:
	:	:	:	:
	:	:	:	:
Qy	305	ATREDFLNPDV-----	LDIFISSYQKVFQVVESTRPGKKVWLGETSSAYGGAPLLSDTFA	360
	:	:	:	:
	:	:	:	:
	:	:	:	:
Db	312	-----NPKDLLLHDDLLISSIVATL-----	-----KSRFIGNLNI-----GTGQLSSITLK	351
	:	:	:	:
	:	:	:	:
	:	:	:	:
Qy	361	AGFMWLDLKLGLSA-----	RMGIEVNVNRQVFFGAGNYHLVDENFDPDLYWLSLIFPKVLUG	415
	:	:	:	:
	:	:	:	:
	:	:	:	:
Db	352	IAEMIRDELGSMMTCOOIEVNTTEVASIAMNYGRAN--HVLQ-----	-----WEPVITFFE--QG	400
	:	:	:	:
	:	:	:	:
	:	:	:	:

```

Qy 416 TKVLMSV 423
Db 401 LKSLHQI 408

RESULT 11
E91031
probable outer membrane protein EC93221 [imported] - Escherichia coli (strain O157:H7, st
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 05-Oct-2004
C:Accession: E91031
C:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, I.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E91031
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-879 <HAY>
A:Cross-references: UNIPROT:Q8XCP4; UNIPARC:UPI0000D0453; GB:BA000007; PIDN:BAB36644.1;
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: EC93221
C:Superfamily: outer membrane usher protein fimD

Query Match 3.7%; Score 105.5; DB 2; Length 879;
Best Local Similarity 19.9%; Pred. No. 13;
Matches 129; Conservative 69; Mismatches 208; Indels 241; Gaps 33;

Qy 52 VPSFSLSVTIDANLATDPRFLILGLSPKRLTARGLSPAYLRFGGTKTDFLI FDPKKEST 111
Db 20 MSGSYVNAWAENEIQFDSRFLLEKGDTKI-DLKRFSQGYVPEG--KYNLQVLNKPQPLT 76

Qy 112 FEERSYQSQVNQDICKVGSIPDPVEKL-----RLEWYQEQOLLREHYQKFKNS 163
Db 77 EYDIYVASENDASKTYACLTPELVAQGLKEDVAKLQWHDGKCLPGQLE----- 130

Qy 164 TYSRSSVDLVTFANCGLDIFGL-NALLRTADLQWNSN-----AQLLLDYC----- 211
Db 131 -----GDIK---ADLSQALVISLPQAYLEYTDINWPPSRWDDGISGLIADYSITAT 182

Qy 212 -----SSKGNYI-SWEI-GNEPNSFL---KKADIFINGSQ----- 241
Db 183 RHEENGGDSDNEISGNGTVGVNLGAWLRLADWQTDYLSKSNDDVDVINGDDTQKNWESR 242

Qy 242 -----LGEDFTQLHKLKSTF-----KNAKLYGDPV 268
Db 243 YVAVRALPSLKAKGLGELY-----LNSDIFDGNVYVGGSISTDQMLPPNLRGVAPDI 296

Qy 269 GQPRKTKAKMLKSLKAGGEVI-----DSVTHHYLYLNGRTATREDFLN 312
Db 297 SGVAHTTAKVTVSQ---GRVIYETQVPAGFPRIQDLGDSV-----SGTLHRIEON 346

Qy 313 PDVLIDIFISSQVKQFQVVESTRPK---KWL-----GETS----- 345
Db 347 GQVQEQYDINTASMPF---LTRPGQVRYKLMWGRPQEWGHVHGVEGFFSGGEASWGIANGW 402

Qy 346 SAYGGGAPLLSD-----TFAA-----GFWWLDKL-----GLSAR 374
Db 403 SLYGGA---LADEHYQSAALGVGRDLSVFGAVAFDITSHSTRLDKETA YGKSLDGNSPR 459

Qy 375 MGI-----EVVMRQVFCAGNYHLVDENFDPLPDYWL SILLFKKL VGT---KVLMA SVQGS 426
Db 460 LSVSKDFDELNSRVTFAG---YRFSEENFTMTSEY-LDASDSEMVRTGNDKEMYTATYNO 515

Qy 427 KRRKRLRVYLHCTNDNPRYKEGDITLYAI-----NLHNVTK----- 462
Db 516 NFRDAGVSVLYNTRYTHRYTWDRDEQTNVYMLSHVFNLGSI RNMSISMTGYRVEYDQADK 575

Qy 463 --YLRLPYPSFNKQVDKYLRPLRPLGPHGLLSKSVQLNGLT LKQVDDOT 507
Db 576 GVYISLGMWGDSTISY---NGNYGSGDSSOVG---YFSRVDDAT 616

```

RESULT 12

S00652  
phosphoribosylamine-glycine ligase (EC 6.3.4.13) - fission yeast (Schizosaccharomyces pombe)  
N;Alternate names: AIRase; aminimidazole ribotide synthetase; GARSase; glycineamide rib  
N;Contains: phosphoribosylamine-glycine ligase (EC 6.3.4.13); phosphoribosylformylglycin  
C;Species: Schizosaccharomyces pombe  
C;Date: 07-Sep-1990 #sequence\_revision 28-Oct-1994 #text\_change 09-Jul-2004  
C;Accession: S00652; T40496; T40422  
R;McKenzie, R.; Schuchert, P.; Kilbey, B.

Curr. Genet. 12, 591-597, 1987  
A;Title: Sequence of the bifunctional ade1 gene in the purine biosynthetic pathway of th  
A;Reference number: S00652; MUID:89003164; PMID:3502942  
A;Accession: S00652

A;Molecule type: DNA  
A;Residues: 1-788 <MCK>  
A;Cross-references: UNIPROT:P20772; UNIPARC:UPI0000132A3F; EMBL:X06601; NID:94903; PIDN:  
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lauber, J.; Hilbert, H.; Duesterhoeft, A.  
submitted to the EMBL Data Library, February 1998  
A;Reference number: Z21910  
A;Accession: T40496

A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-788 <WOO>  
A;Cross-references: UNIPARC:UPI0000132A3F; EMBL:AL021730; PIDN:CAAL6823.1; GSPDB:GN00067  
A;Experimental source: strain 972h-; cosmid c4C3  
R;Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, March 1999  
A;Reference number: Z21928  
A;Accession: T40422

A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 604-788 <SEB>  
A;Cross-references: UNIPARC:UPI000016908D; EMBL:AL035655; PIDN:CAB38600.1; GSPDB:GN00067  
A;Experimental source: strain 972h-; cosmid c405  
C;Genetics:

A;Map position: 2  
A;Gene: ADE1; SPDB:SPBC405.01  
A;Superfamily: Saccharomyces cerevisiae ADE5 multifunctional protein; phosphoribosylamin  
C;Keywords: cyclo-ligase; purine nucleotide biosynthesis  
F;5-425/Domain: phosphoribosylamine-glycine ligase homology <PGL>  
F;439-767/Domain: phosphoribosylformylglycinamide cyclo-ligase homology <PFCL>

Query Match 3.7%; Score 104.5; DB 1; Length 788;  
Best Local Similarity 27.7%; Pred. No. 13;  
Matches 70; Conservative 36; Mismatches 114; Indels 33; Gaps 11;

QY 297 HYVINGRTATRE--DFLNPDV-LDIFISSVQKVFQVVEST-RPGKKVWLGETSSAY---- 348  
DB 424 HHALNPKRTREILTYENSGVSDNGNEFQRIKDLVKSTRPGADADIGCGGIFDLKQ 483  
QY 349 -GGGAPLL-SDTPAAGFMWLDKLGLSAR--MGIEVNRQVFFGAGNYHLVDENFDPL--P 402  
DB 484 AGWNDPLLVSATDGVGSKLLIALSLNKHDTVGIDLVMNV-----NDLVVQGAEPILFL 537  
QY 403 DYWLSLLFKLVCTKVMASVQSGSKRKLRYVILHCTNDNPRYKEGDLTYALNHLNVTK 462  
DB 538 DYFATGSLDLKVTFSFVEGVKKGQAGCALVGGETSEMPGLYHDGHDYDANGTSVGAVSR 597  
QY 463 YLRLPYPFENKQVDKYLRLPLGPHGLLSKSVQLNGLTL--KMVD-----DQTLPLPMEKPL 516  
DB 598 DDLPLPESFGKDILL-----GLASDGVHNSGYSLVRKIVEYSDLEYSVCPCWDKNV 650  
QY 517 RGGSSLGLPAFSY 529  
DB 651 RLGDSSLIPRIY 663

RESULT 13

F70411  
adenylosuccinate synthetase - Aquifex aeolicus  
C;Species: Aquifex aeolicus

C;Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004  
C;Accession: F70411  
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove  
V.

Nature 392, 353-358, 1998  
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A;Reference number: A70300; MUID:98196666; PMID:9537320  
A;Accession: F70411

A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-432 <AQF>

A;Cross-references: UNIPROT:O67321; UNIPARC:UPI00000565A2; GB:AE000733; NID:g2983720; PII  
A;Experimental source: strain VF5  
C;Genetics:

A;Gene: purA  
C;Superfamily: adenylosuccinate synthase

Query Match 3.7%; Score 104; DB 2; Length 432;  
Best Local Similarity 23.9%; Pred. No. 5.9;  
Matches 96; Conservative 39; Mismatches 128; Indels 138; Gaps 22;

QY 15 LLLGLGLGSPGALPRPAQADVVDLD-----PFTQEPHLVSPS 55  
DB 51 ILHLLPTGILHEHVKGVIAGQM-VVDLEVLHKEVNLEEKGIYVKERLIFSDRAHLVMPY 109

QY 56 FLSVTIDANLATDPRFLILLGSPK--LRTLARGLSPAYL-RFGGKTDTDLIFDPKKESTF 112  
DB 110 H-----KLLDSLFEKKKGIGTTLRGIGPAYMPKYG--RKGIIRISDLKDKRF 154

QY 113 EERSYQSQVNQDICKYGSIPPDVEEK-----LRLEWPYOEQLLREHYQKPKPNSTY 165  
DB 155 ----YTLLEDNLDPVK-----NICEKVFCEKFDLDINQIYEEQL----RYFEFKEKV- 199

QY 166 SRSDVDVLYTFANCGLDLIFGLNALLRTADL----QWNSNAQLLLDYCSSKGYNISWE 221  
DB 200 ----VDLLRFFNTQKGSVLFEAGQGTLLDMDMGTYPPYVTSSNASAL-----GLSNG 246

QY 222 LGNEPNSFLKKADIFING-----SQL-GEDFIQLHLKLRKSTFKNAKLYG 265  
DB 247 TGMPPKYF---SDAFFLGVAKAYATRVGEGPPFPTELKGEKEKREL-----GGYIG 295

QY 266 PDVQGPQR---KTAKMLKSLKAGGEVIDSVTHHHYLLNGRTATREDPLNP----- 313  
DB 296 STTGRPRRCGWLDLVALKYAVQVNG-----LDGFVITKLDVLDTFDEVKVCVA 343

QY 314 -----DVLDIFISVQKVFQV--VESTRPCKKXWLGESSA 347  
DB 344 YELDGEVIDYFPASYSSELIRKVPYKTLKG---WKKSTKGA 381

RESULT 14

D97065

transketolase [imported] - Clostridium acetobutylicum  
C;Species: Clostridium acetobutylicum

C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C;Accession: D97065

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001

A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A;Reference number: A96900; MUID:21359325; PMID:21359325

A;Accession: D97065

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-796 <KUR>

A;Cross-references: UNIPROT:Q97JE3; UNIPARC:UPI0000131969; GB:AE001437; PIDN:AAK79311.1;

A;Experimental source: Clostridium acetobutylicum ATCC824

C;Genetics:

A;Gene: CAC1343

C;Superfamily: phosphoketolase

Query Match 3.7%; Score 104; DB 2; Length 796;  
Best Local Similarity 21.6%; Pred. No. 15;



Matches	77;	Conservative	51;	Mismatches	114;	Indels	114;	Gaps	18;
Qy	27	GALPREAQAQDVVDLDFFTQEPHLVSPGFLSVTIDA---NLATDPRF-LILGSPKLT 82							
Db	238	GWKPYFVEGEDPETMHLMAETLDIVTEELINIQKNARENDCSRPKWPMIVLRTEK--- 294							
Qy	83	LARGLSPAYLRFQGTKTDFLIFDPKKESTPEERSYQSQVNQDICKYGSIPPDVEEKRL 142							
Db	295	-----GWTGPKFV-----DGVNPEGSRFAHQVPLAVDRYHTENLDQLE----- 332							
Qy	143	EW--PYQEOILLRHHYQ--KKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLQ 198							
Db	333	EWLKSYPPELFDENYRLIPELEBLTPGKNKMAANLHAN--GGL-----LLRELTPDPR 386							
Qy	199	WNSNAQLLDYSSKGYNISWELGNPNFSLKKADIFINGSQIGEDFIQLHKLRL----- 254							
Db	387	-----DYA-----VDVPTPGSTVKQDMIELGKYVRDVVK 415							
Qy	255	-KSTFKNAKLYGPD-----VGQPRKTKAMLK---SFLKAGEVIDSVTWHH-- 297							
Db	416	LNEDTRNFRIFGPDETMSNRLMAVFEQTKQWLSEIKEPNDEFSLNDGRIVDSMLSEHLC 475							
Qy	298	-----YLLNGRTATREDFLNPDVLDIFISSVQKVFQWVES--TRPGKKVWLGETS 345							
Db	476	EGWLEGYLLTGRHG-----FFASYEAFRLIVDSMITQHGK--WLKVTS 516							
RESULT 15									
A11489									
probable peptidoglycan bound protein (LPXTG motif) lin0457 [imported] - Listeria innocua									
C:Species: Listeria innocua									
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004									
C:Accession: A11489									
.; Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.									
Science 294, 849-852, 2001									
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maier, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.									
A:Reference number: AB1077; MUID:21537279; PMID:11679669									
A:Accession: A11489									
A>Status: preliminary									
A:Molecule type: DNA									
A:Residues: 1-2013 <GLA>									
A:Cross-references: UNIPROT:Q92EK2; UNIPARC:UPI00000CC237; GB:AL592022; PIDN:CAC95689.1;									
A:Experimental source: strain Clip11262									
C:Genetics:									
A:Gene: lin0457									
Query Match 3.7%; Score 104; DB 2; Length 2013;									
Best Local Similarity 20.6%; Pred. No. 60;									
Matches	95;	Conservative	64;	Mismatches	161;	Indels	142;	Gaps	21;
Qy	5	SKPALPPPLMLLLGLPLSPGALPRPAQADVVDLD-----FFT 45							
Db	256	AKQGVPMNLKATL---SGENSAGATYTPAEKTTVNLEENSSNLDSPITAGDNSWAFSM 312							
Qy	46	QEPHLVSPFLSV---TIDANLATDPRFLILLGSPKLTLRG-----LSPAYLR 93							
Db	313	KELAFSLKPGGYTTIQWPEIQKSSSENKSFKNL----KLEFLKENGDDIISVNTADPYVIR 368							
Qy	94	FG-----GKTFDLFPDKKESTPEERSYQSQVNQDICKYGSIPPDVEEKRL 143							
Db	369	FGEYPWSQLSTVNGKANVLNDEKQ-----VVEYGPINANIYQRIQVS 412							
Qy	144	WP-----YQQLLLRHHYQKFKNSTYSRSSVDVLYTFANCSGLDL 184							
Db	413	MAAKIPADAVKGTGYTGVNVYDSDLVTSIKITVETDTSATSAVDSKVSITSISGDV 472							
Qy	185	IFGLNALLRTADLQWN-----SSNA-----QLLLDYCSSKGYNI--SWELGNPNFSLKKAD 234							
Db	473	-----LEWGFMPRISSAAPGVNDLEIVAPIPKIGIKVLSYIINNNSMASMKKLE 520							

Qy	235	IFING-----SOLGEDFIQLHKLRL-----KSTFKNAKLYGPDVGOPRRKTKAKWLS 281						
Db	521	YYQNGKWYSMAPOQTSSGWDFSKIDQSVNRIEKCLKLTSRDGIINDKMPPTHTGTHRMQNT 580						
Qy	282	FLKAGGEVI---DSVTWHYYLNGRTATREDFLNPDVLDIFISSVQKVFQWVEST--RPGK 337						
Db	581	GVKAGESFTLOPESIT---YTDSDKTS-----KAIDTTASSYEKKVQVVEKTSTPAK 629						
Qy	338	---KWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMG 376						
Db	630	INGDVFLSSTAGIYKG--FESTIFFNG---DKIAQSVRLG 665						

Search completed: June 5, 2006, 12:21:41  
Job time : 21.0785 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:21 ; Search time 134.728 Seconds  
(without alignments)  
3728.138 Million cell updates/sec

Title: US-10-645-659A-4  
Perfect score: 2841  
Sequence: 1 MLRSKPALPPPLMLLLGP.....LPASYSFFVIRNAKVAACI 543

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 7.2.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2841	100.0	543	1	HPSE HUMAN
2	2285	80.4	545	1	HPSE BOVIN
3	2276	80.1	574	2	Q333X9 GRODE
4	2275	80.1	574	2	Q333X8 SPALAX GOLA
5	2263	79.7	574	2	Q333X7 GRODE
6	2262	79.6	574	2	Q333X6 SPALAX JUDA
7	2170	76.4	535	1	HPSE MOUSE
8	2154	75.8	558	2	Q333X5 SPALAX JUDA
9	2138	75.3	536	1	HPSE RAT
10	1648	58.0	523	1	HPSE CHICK
11	1320	46.5	533	2	Q4SYF6 TETNG
12	1150.5	40.5	592	1	HPSE2 HUMAN
13	1150.5	40.5	592	2	Q2MLH9 HUMAN
14	1036.5	36.5	597	2	Q4TB80 TETNG
15	742.5	26.1	255	2	Q4TGC8 TETNG
16	699	24.6	535	2	Q8TI08_BOMMO
17	417	14.7	543	1	HPSE1 ARATH
18	400	14.1	559	2	Q89F99 BRAJA
19	388	13.7	526	2	Q5SNA6 ORYSA
20	379	13.3	541	2	Q69II6 ORYSA
21	377	13.3	527	2	Q9LRC8 SCUBA
22	368	13.0	536	1	HPSE3 ARATH
23	365	12.8	537	2	Q70YJ3 HORVU
24	359.5	12.7	539	2	Q2QN56 ORYSA
25	354	12.5	401	2	Q30324 ARATH
26	353.5	12.4	539	1	HPSE2 ARATH
27	350.5	12.3	529	2	Q6ZUE2 ORYSA
28	323.5	11.4	516	2	Q447R5 SOLUS
29	292	10.3	537	2	Q43S03 SOLUS
30	276.5	9.7	506	2	Q37070 SPHAR
31	245.5	8.6	382	2	Q3E8P7_ARATH

32	156	5.5	935	2	Q9VE79 DROME	Q9ve79 drosophila
33	141	5.0	559	2	Q7SF80_NEUCR	Q7sf80 neurospora
34	140.5	4.9	536	2	Q2UDS9_ASFOR	Q2uds9 aspergillus
35	138.5	4.9	463	2	Q63T97 BURPS	Q63t97 burkholderi
36	138.5	4.9	670	2	Q3JTG0 BURP1	Q3jtg0 burkholderi
37	138	4.9	1128	2	Q5TT65 ANOGA	Q5tt65 anopheles g
38	136.5	4.8	795	2	Q2ZPT8 SHEPU	Q2zpt8 shewanella
39	135.5	4.8	795	2	Q2X712_9GAMM	Q2x712 shewanella
40	134	4.7	510	2	Q2U0T3_ASFOR	Q2u0t3 aspergillus
41	126.5	4.5	493	2	Q2HK01_THEAC	Q2hk01 thermoplasma
42	125.5	4.4	665	2	Q55IC3 CRYNE	Q55ic3 cryptosporidium
43	125.5	4.4	665	2	Q5K7V9 CRYNE	Q5k7v9 cryptosporidium
44	124	4.4	634	2	Q5N1L7_9BACT	Q5n1l7 uncultured
45	124	4.4	765	2	Q4P0C9_USTMA	Q4p0c9 ustilago ma

ALIGNMENTS

RESULT 1  
HPSE\_HUMAN STANDARD; PRT; 543 AA.  
ID HPSE\_HUMAN  
AC Q9Y251; Q53GE5; Q9UL39;  
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
DT 01-NOV-1999, sequence version 1.  
DT 07-FEB-2006, entry version 27.  
DE Heparanase precursor (EC 3.2.-.-) (Heparanase-1) (Hpal) (Endo-glucuronidase) [Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].  
GN Name=HPSE; Synonyms=HEP, HPA, HPAL, HPR1, HPSE1, HSE1;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.  
RC TISSUE=Placenta;  
RX MEDLINE=99335379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;  
RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C., Seddon A.P., Giorgio N.A., Bohlen P.;  
RT "Cloning and functional expression of a human heparanase gene.";  
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN SEQUENCE OF 158-168; 326-337 AND 447-491.  
RC TISSUE=Embryonic fibroblast;  
RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;  
RA Toyoshima M., Nakajima M.;  
RT "Human heparanase. Purification, characterization, cloning, and expression.";  
RL J. Biol. Chem. 274:24153-24160(1999).  
RN [3]  
RP NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.  
RX PubMed=10395325; DOI=10.1038/10518;  
RA Vlodaysky I., Friedmann Y., Elkin M., Aingorn H., Atzmon R., Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I., Spector L., Becker I.;  
RT "Mammalian heparanase: gene cloning, expression and function in tumor progression and metastasis.";  
RL Nat. Med. 5:793-802(1999).  
RN [4]  
RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE OF 158-174; 263-272; 326-337; 433-436; 438-443; 466-468 AND 478-483.  
RC TISSUE=Placenta;  
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;  
RA Hullett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J., Parish C.R.;  
RT "Cloning of mammalian heparanase, an important enzyme in tumor invasion and metastasis.";  
RL Nat. Med. 5:803-809(1999).  
RN [5]  
RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE

- RP SPECIFICITY.
- RA TISSUE=Placenta;
- RX MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;
- RT Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;
- RA "Heparanase expression in invasive trophoblasts and acute vascular
- RT damage.";
- RL Glycobiology 10:467-475 (2000).
- RN [6]
- RN NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.
- RP PubMed=11547900; DOI=10.1023/A:1011375624902;
- RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O.,
- RA Aviv A., Elkin M., Pecker I., Peretz T., Vlodavsky I.;
- RT "Molecular properties and involvement of heparanase in cancer
- RT progression and mammary neoplasia genesis.";
- RL J. Mammary Gland Biol. Morphogenesis 6:311-322 (2001).
- RN [7]
- RN NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163,
- RP SUBUNITS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.
- RC TISSUE=Placenta;
- RX PubMed=12713442; DOI=10.1042/BJ20030318;
- RA McKenzie E., Young K., Hircok M., Bennett J., Bhaman M., Felix R.,
- RA Turner P., Stamps A., McWilliam D., Saville G., Ng S., Mason S.,
- RA Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M.,
- RA Farekh R., Stuberfield C.;
- RT "Biochemical characterization of the active heterodimer form of human
- RT heparanase (Hpa1) protein expressed in insect cells.";
- RL Biochem. J. 373:423-435 (2003).
- RN [8]
- RN NUCLEOTIDE SEQUENCE [MRNA].
- RP Pinal M.A., Semedo P.;
- RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,
- RA Tanaka A., Yokoyama S.;
- RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
- RN [10]
- RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
- RC TISSUE=Pancreas;
- RX MEDLINE=2338257; PubMed=12477932; DOI=10.1073/pnas.242603899;
- RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
- RA Lausner R.D., Collins F.S., Wagner L., Shennen C.M., Schuler G.D.,
- RA Altschul S.P., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
- RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
- RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
- RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
- RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
- RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
- RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
- RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
- RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
- RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
- RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
- RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
- RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
- RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
- RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
- RT "Generation and initial analysis of more than 15,000 full-length human
- RT and mouse cDNA sequences";
- RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
- RN [11]
- RN MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.
- RX PubMed=11123890; DOI=10.1021/bi002080p;
- RA Hulett M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,
- RA Gready J.E., Parish C.R.;
- RT "Identification of active-site residues of the pro-metastatic
- RT endoglycosidase heparanase.";
- RL Biochemistry 39:15659-15667 (2000).
- RN [12]
- RP N-GLYCOSYLATION, AND MUTAGENESIS OF ASN-162; ASN-178; ASN-200;
- RP ASN-217; ASN-238 AND ASN-459.
- RX PubMed=14573609; DOI=10.1074/jbc.M300541200;
- RA Smizu S., Ishida K., Wierzbicka M.K., Osada H.;
- RT "Secretion of heparanase protein is regulated by glycosylation in
- RT human tumor cell lines.";
- RL J. Biol. Chem. 279:2697-2703 (2004).
- RN [13]
- RN SUBCELLULAR LOCATION.
- RP PubMed=15292202; DOI=10.1074/jbc.M402131200;
- RA Gingis-Velitski S., Zetser A., Kaplan V., Ben-Zaken O., Cohen E.,
- RA Levy-Adam F., Bashenko Y., Flugelman M.Y., Vlodavsky I., Ilan N.;
- RT "Heparanase uptake is mediated by cell membrane heparan sulfate
- RT proteoglycans.";
- RL J. Biol. Chem. 279:44084-44092 (2004).
- RN [14]
- RN BIOPHYSICOCHEMICAL PROPERTIES, PROCESSING, AND SUBCELLULAR LOCATION.
- RP PubMed=15848168; DOI=10.1016/j.febslet.2005.03.030;
- RA Cohen E., Atzmon R., Vlodavsky I., Ilan N.;
- RT "Heparanase processing by lysosomal/endosomal protein preparation.";
- RL FEBS Lett. 579:2334-2338 (2005).
- RN [15]
- RN SUBCELLULAR LOCATION, PROCESSING, AND MUTAGENESIS OF TYR-156.
- RP PubMed=15659389; DOI=10.1074/jbc.M413370200;
- RA Abboud-Jarrous G., Rangini-Guetta Z., Aingorn H., Atzmon R.,
- RA Elgavish S., Peretz T., Vlodavsky I.;
- RT "Site-directed mutagenesis, proteolytic cleavage, and activation of
- RT human proheparanase.";
- RL J. Biol. Chem. 280:13568-13575 (2005).
- RN [16]
- RN DOMAINS, AND MUTAGENESIS OF LYS-158 AND LYS-161.
- RP PubMed=15760902; DOI=10.1074/jbc.M414546200;
- RA Levy-Adam F., Abboud-Jarrous G., Guerrini M., Beccati D.,
- RA Vlodavsky I., Ilan N.;
- RT "Identification and characterization of heparin/heparan sulfate
- RT binding domains of the endoglycosidase heparanase.";
- RL J. Biol. Chem. 280:20457-20466 (2005).
- RN [17]
- RN VARIANT SER-260.
- RP PubMed=15334672;
- RA Chen X.P., Liu Y.B., Rui J., Peng S.Y., Peng C.H., Zhou Z.Y.,
- RA Shi L.H., Shen H.W., Xu B.;
- RT "Heparanase mRNA expression and point mutation in hepatocellular
- RT carcinoma.";
- RL World J. Gastroenterol. 10:2795-2799 (2004).
- CC -I- FUNCTION: Endoglycosidase which is a cell surface and
- CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
- CC proteoglycans (HSPGs) into heparan sulfate side chains and core
- CC and tumor cell lines. Due to its contribution to metastasis and
- CC angiogenesis, it is considered to be a potential target for anti-
- CC cancer therapies.
- CC -I- ENZYME REGULATION: Inhibited by EDTA, laminarin sulfate and, to a
- CC lower extent, by heparin and sulfamin and activated by calcium and
- CC magnesium (by similarity).
- CC -I- BIOPHYSICOCHEMICAL PROPERTIES:
- CC pH dependence:
- CC Optimum pH is 4-6;
- CC -I- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
- CC subunits, the proteolytic products.
- CC -I- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
- CC Secreted, internalised and transferred to late endosomes/lysosomes
- CC as a proheparanase. In lysosomes, it is processed into the active
- CC form, the heparanase. The uptake or internalisation of
- CC proheparanase is mediated by HSPGs. Heparin appears to be a
- CC competitor and retain proheparanase in the extracellular medium.
- CC -I- TISSUE SPECIFICITY: Highly expressed in placenta and spleen and
- CC weakly expressed in lymph node, thymus, peripheral blood
- CC leukocytes, bone marrow, endothelial cells, fetal liver and tumor
- CC tissues
- CC -I- PM: Proteolytically processed. The cleavage of the 65 kDa form
- CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
- CC product. The active form, the 8/50 kDa heterodimer, is resistant
- CC to degradation. Complete removal of the linker peptide appears to
- CC be a prerequisite to the complete activation of the enzyme.

```
CC -1- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility.

Query Match 100.0%; Score 2841; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 2e-205;
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRSKPALPPPPMLLLGLPLSPALPRPAQDVVDLDFPTQPLHLVSPFLSVT 60
DB 1 MLRSKPALPPPPMLLLGLPLSPALPRPAQDVVDLDFPTQPLHLVSPFLSVT 60
QY 61 IDANLATDPRFLLILGSPKLTTLARGLSPAYLRFGGTKTDFLFDPKKESTFEERSYQWS 120
DB 61 IDANLATDPRFLLILGSPKLTTLARGLSPAYLRFGGTKTDFLFDPKKESTFEERSYQWS 120
QY 121 QVNQDICKYGSIPDPVEEKRLWEPYQEQLLREHYOKKFNSTYSRSSVDVLYTFANCS 180
DB 121 QVNQDICKYGSIPDPVEEKRLWEPYQEQLLREHYOKKFNSTYSRSSVDVLYTFANCS 180
QY 181 GLDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240
DB 181 GLDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFINGS 240
QY 241 QLCGEDFIQLHLKLRKSTFKNAKLYGPDVGQPRKTAAMLKSFLLKAGGEVIDSVTWHYYL 300
DB 241 QLCGEDFIQLHLKLRKSTFKNAKLYGPDVGQPRKTAAMLKSFLLKAGGEVIDSVTWHYYL 300
QY 301 NGRTATREDFLNPDLVDIFISSVQKVFQVVESTRPGKKVWLGETSSAYGCGAPLLSDTFA 360
DB 301 NGRTATREDFLNPDLVDIFISSVQKVFQVVESTRPGKKVWLGETSSAYGCGAPLLSDTFA 360
QY 361 AGFMWLDKLGSLARMGIEVVMQVFFGAGNYHLVDENFDPLPDYWLSSLFPKLVGTKVL 420
DB 361 AGFMWLDKLGSLARMGIEVVMQVFFGAGNYHLVDENFDPLPDYWLSSLFPKLVGTKVL 420
QY 421 ASVQGSKRKRRLRYLHCTNTDNPYKEGDLTLVAINLHNVTYKLRLPYPSPNKQVDKYL 480
DB 421 ASVQGSKRKRRLRYLHCTNTDNPYKEGDLTLVAINLHNVTYKLRLPYPSPNKQVDKYL 480
QY 481 RPLGPHGLLSKSVQLNGLTLKMVDDQTLPLMEKPLRPGSSLGIPAFSFFVIRNAKVA 540
DB 481 RPLGPHGLLSKSVQLNGLTLKMVDDQTLPLMEKPLRPGSSLGIPAFSFFVIRNAKVA 540
QY 541 ACI 543
DB 541 ACI 543

RESULT 2
ID HPSE BOVIN STANDARD; PRT; 545 AA.
AC Q9MY0;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-JUN-2001, sequence version 2.
DT 07-MAR-2006, entry version 15.
DE Heparanase precursor (EC 3.2.-.-) [Contains: Heparanase 8 kDa subunit;
DE Heparanase 50 kDa subunit].
CN Name:HPSE;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RC TISSUE=Placenta;
RX MEDLINE=21176669; PubMed=11277877;
RA Kizaki K., Nakano H., Nakano H., Takahashi T., Imai K., Hashizume K.;
RT "Expression of heparanase mRNA in bovine placenta during gestation.";
RL Reproduction 121:573-580(2001).
CC -1- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
```

```
CC proteoglycans. Also implicated in the extravasation of leukocytes
CC and tumor cell lines. Contributes to metastasis and angiogenesis
CC (By similarity).
CC -1- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower
CC extent, by heparin, sulfamin and EDTA. Activated by calcium and
CC magnesium (By similarity).
CC -1- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa
CC subunits, the proteolytic products (By similarity).
CC -1- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.
CC Secreted, internalised and transferred to late endosomes/lysosomes
CC as a proheparanase. In lysosomes, it is processed into the active
CC form, the heparanase. The uptake or internalisation of
CC proheparanase is mediated by HSPGs. Heparin appears to be a
CC competitor and retain proheparanase in the extracellular medium
CC (By similarity).
CC -1- TISSUE SPECIFICITY: Highly expressed in placenta and weakly in the
CC kidney, lung, spleen and uterus.
CC -1- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC product. The active form, the 8/50 kDa heterodimer, is resistant
CC to degradation. Complete removal of the linker peptide appears to
CC be a prerequisite to the complete activation of the enzyme (By
CC similarity).
CC -1- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility (By similarity).
CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonCommercial License
CC
CC EMBL: AF281160; AAF87301.2; -; mRNA.
CC InterPro: IPR005199; Glyco_hydro_79_N.
CC Pfam: PF03662; Glyco_hydro_79n; 1.
CC KX Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;
CC SIGNAL.
CC FT SIGNAL 1 37 By similarity.
CC FT CHAIN 38 111 Heparanase 8 kDa subunit (By similarity).
CC FT PROPEP 112 159 /FTID=PRO_0000042256.
CC FT CHAIN 160 545 /FTID=PRO_0000042257.
CC FT CHAIN 160 545 Heparanase 50 kDa subunit (By
CC similarity).
CC FT /FTID=PRO_0000042258.
CC FT REGION 160 164 Heparin/HS-binding (Potential).
CC FT REGION 272 282 Heparin/HS-binding (Potential).
CC FT ACT_SITE 227 227 Proton donor (Potential).
CC FT ACT_SITE 345 345 Nucleophile (Potential).
CC FT CARBOHYD 164 164 N-linked (GLCNAC...) (Potential).
CC FT CARBOHYD 219 219 N-linked (GLCNAC...) (Potential).
CC FT CARBOHYD 461 461 N-linked (GLCNAC...) (Potential).
CC SQ SEQUENCE 545 AA; 61077 MW; FAC4BDFD855B933 CRC64;

Query Match 80.4%; Score 2285; DB 1; Length 545;
Best Local Similarity 80.0%; Pred. No. 1.9e-163;
Matches 436; Conservative 34; Mismatches 73; Indels 2; Gaps 1;

QY 1 MLRSKPALPPPPMLLLGLPLSPALPRPAQDVVDLDFPTQPLHLVSPFLSVT 58
DB 1 MLRSKPALPPPPMLLLGLPLSPALPRPAQDVVDLDFPTQPLHLVSPFLSVT 60
QY 59 VTIDANLATDPRFLLILGSPKLTTLARGLSPAYLRFGGTKTDFLFDPKKESTFEERSYQWS 118
DB 61 FTIDANLATDPRFLLILGSPKLTTLARGLSPAYLRFGGTKTDFLFDPKKESTFEERSYQWS 120
QY 119 QSVNQDICKYGSIPDPVEEKRLWEPYQEQLLREHYOKKFNSTYSRSSVDVLYTFAN 178
DB 121 LSQSNQDICKYGSIPDPVEEKRLWEPYQEQLLREHYOKKFNSTYSRSSVDVLYTFAN 180
QY 179 CSGDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFIN 238
DB 181 CSGDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYNISWELGNEPNSFLKKADIFIN 240
QY 239 GSQLGEDFIQLHLKLRKSTFKNAKLYGPDVGQPRKTAAMLKSFLLKAGGEVIDSVTWHYY 298
```



```

QY 60 TIDANLATDPRFLILGSPKLTARGLSPAYLRFGGTKTDFLIDFPKKESTFEERSYQ 119
DB 91 TIDANLATDPRFLILGSPKLTARGLSPAYLRFGGTKTDFLIDFPKKESTFEERSYQ 150
QY 120 SOVNQDICKYGSIPPDVEEKRLRLEWYQEQQLLLREHYQKFKNSTYSRSSVDMLYTFANC 179
DB 151 SOVNHDICRSGAIPAVVVRRLQVEWPFQEQQLLLREHYQKFKNSTYSRSSVDMLYTFANC 210
QY 180 SGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFING 239
DB 211 SGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNEPNSFWKKAHISIDG 270
QY 240 SOLGEDFIQLHKLKRSKSTFNKAKLYGPDVGOPRRKTAKMLKSLFKAGGEVIDSVTWHYY 299
DB 271 LQAGEDYIELHKLKRSKSTFNKAKLYGPDVGOPRRKTAKMLKSLFKAGGEVIDSVTWHYY 330
QY 300 LNGRTATREDFLNPDVLDIFISSVQKVFQVVESTPRGKVKWLGETSSAYGGGAPLLSDTF 359
DB 331 LNGRIATKEDFLSPDVLDTFILSVQKILQVVEETPRGKVKWLGETSSAYGGGAPLLSNTF 390
QY 360 AAGFMWLDKLGSLARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKLVGTQVL 419
DB 391 AAGFMWLDKLGSLARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKLVGTQVL 450
QY 420 MASVQGSKRERKRLVYLHCTNTDNRYKEGDLTLVAINLHNTYKYLRLPYFPNSKQVDKYL 479
DB 451 MARVKGDRSKRLVYLHCTNTDNRYKEGDLTLVAINLHNTYKYLRLPYFPNSKQVDKYL 510
QY 480 LRPLGPHGLLSKSVQNLGLTLKMWDDQTLPLMEKPLRPGSSGLPAPFSYFFVIRNAKV 539
DB 511 VKPLGPGGLLSKSVQNLGLTLKMWDDQTLPLMEKPLRPGSSGLPAPFSYFFVIRNAKV 570
QY 540 AACI 543
DB 571 AACL 574

```

```

RESULT 5
Q333X7_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X7;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax carmeli.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164324;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avioli A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085492; CAJ30019.1; -; mRNA.
SQ SEQUENCE 574 AA; 64459 MW; 9FD19DCBBD99DE CRC64;

```

```

Query Match 79.7%; Score 2263; DB 2; Length 574;
Best Local Similarity 80.1%; Pred. No. 9.4e-162;
Matches 436; Conservative 42; Mismatches 60; Indels 6; Gaps 2;
QY 4 RSKPALPPPLMLLLL-----GPLGFLSGALPRPAQAQDVVDLFFQEPHLHVSPLSV 59
DB 33 RCQGPGEPLRLSLLLMLWGLSPLVQCIL--AAQADVVVELEFSTQRLHLVSPSPLSI 90

```

```

QY 60 TIDANLATDPRFLILGSPKLTARGLSPAYLRFGGTKTDFLIDFPKKESTFEERSYQ 119
DB 91 TIDANLATDPRFLILGSPKLTARGLSPAYLRFGGTKTDFLIDFPKKESTFEERSYQ 150
QY 120 SOVNQDICKYGSIPPDVEEKRLRLEWYQEQQLLLREHYQKFKNSTYSRSSVDMLYTFANC 179
DB 151 SOVNHDICRSGAIPAVVVRRLQVEWPFQEQQLLLREHYQKFKNSTYSRSSVDMLYTFANC 210
QY 180 SGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNEPNSFLKKADIFING 239
DB 211 SGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYNISWELGNEPNSFWKKAHISIDG 270
QY 240 SOLGEDFIQLHKLKRSKSTFNKAKLYGPDVGOPRRKTAKMLKSLFKAGGEVIDSVTWHYY 299
DB 271 LQAGEDYIELHKLKRSKSTFNKAKLYGPDVGOPRRKTAKMLKSLFKAGGEVIDSVTWHYY 330
QY 300 LNGRTATREDFLNPDVLDIFISSVQKVFQVVESTPRGKVKWLGETSSAYGGGAPLLSDTF 359
DB 331 LNGRIATKEDFLSPDVLDTFILSVQKILQVVEETPRGKVKWLGETSSAYGGGAPLLSNTF 390
QY 360 AAGFMWLDKLGSLARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKLVGTQVL 419
DB 391 AAGFMWLDKLGSLARMGIEVVMRQVFFGAGNYHLVDENFDPLDYWLSLLFKLVGTQVL 450
QY 420 MASVQGSKRERKRLVYLHCTNTDNRYKEGDLTLVAINLHNTYKYLRLPYFPNSKQVDKYL 479
DB 451 MARVKGDRSKRLVYLHCTNTDNRYKEGDLTLVAINLHNTYKYLRLPYFPNSKQVDKYL 510
QY 480 LRPLGPHGLLSKSVQNLGLTLKMWDDQTLPLMEKPLRPGSSGLPAPFSYFFVIRNAKV 539
DB 511 VKPLGPGGLLSKSVQNLGLTLKMWDDQTLPLMEKPLRPGSSGLPAPFSYFFVIRNAKV 570
QY 540 AACI 543
DB 571 AACL 574

```

```

RESULT 6
Q333X6_SPAJD PRELIMINARY; PRT; 574 AA.
AC Q333X6;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax judaei (Blind subterranean mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=134510;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avioli A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085493; CAJ30020.1; -; mRNA.
SQ SEQUENCE 574 AA; 64515 MW; 3AEBB13F07451684 CRC64;

```









QY 180 SGLDIFGLNALLRTADLQWSSNAQLLDYCSSKGVNLSWELGNPNPSFLKKADIFING 239  
Db 211 SGLDIFGLNALLRTADLQWSSNAQLLDYCSSKGVNLSWELGNPNPSFLKKADIFING 270  
QY 240 SOLGEDFIQHLKLLKSTFKNAKLYGPDVQGPQRRKTAQMLKSLKAGGEVIDSVTWHYY 299  
Db 271 LQGEDYIELKLLKSTLKNVLYGPDVQGPQRRKTAQMLKSLKAGGEVIDSVTWHYY 314  
QY 300 LNGRTATREDFENPVDLIFISSVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 359  
Db 315 LNGRTATREDFENPVDLIFISSVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 374  
QY 360 AAGFWLWDLKGLSARNGIEVWQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 419  
Db 375 AAGFWLWDLKGLSARNGIEVWQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 434  
QY 420 MASVQSKRKLRLVYLHCTNTDPRYKQEGDLTYALNHNVTYKLYLPYFNSKQVDKYL 479  
Db 435 MARVQGPDRSKRLVYLHCTNTDPRYKQEGDLTYALNHNVTYKLYLPYFNSKQVDKYL 494  
QY 480 LRPLGPHGLLSKSVQLNGLTLKMVDQDTLPPLMEKPLRPGSSGLPFAFSYFFVIRNAKV 539  
Db 495 VIPLGPGGLLSKSVQLNGLTLKMVDQDTLPPLMEKPLRPGSSGLPFAFSYFFVIRNAKV 554  
QY 540 AACI 543  
Db 555 AACL 558

RESULT 9  
ID HPSE RAT STANDARD; PRT; 536 AA.  
AC Q71RP1; Q90ZF8;  
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
DT 05-JUL-2006, sequence version 1.  
DT 07-MAR-2006, entry version 11.  
DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) (Contains:  
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit).  
GN Name=Hase; Synonyms=Hep;  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridea; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]\_TaxID=10116;  
RP NUCLEOTIDE SEQUENCE [MRNA].  
RC TISSUE=Placenta;  
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;  
RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,  
RA Parish C.R.;  
RT "Cloning of mammalian heparanase, an important enzyme in tumor  
RT invasion and metastasis.";  
RL Nat. Med. 5:803-809(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.  
RX MEDLINE=22194309; PubMed=12077130; DOI=10.1074/jbc.M203282200;  
RA Podyna-Inoue K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;  
RT J. Biol. Chem. 277:32459-32465(2002).  
CC -!- FUNCTION: Endoglycosidase which is a cell surface and  
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
CC proteoglycans. Also implicated in the extravasation of leukocytes  
CC and tumor cell lines. Contributes to metastasis and angiogenesis  
CC (By similarity).  
CC -!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower  
CC extent, by heparin and sulfamin (By similarity). Activated by  
CC calcium and magnesium. Inhibited by EDTA.  
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
CC subunits, the proteolytic products (By similarity).  
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.  
CC Secreted, internalised and transferred to late endosomes/lysosomes  
CC as a heparanase. In lysosomes, it is processed into the active

form, the heparanase. The uptake or internalisation of  
proheparanase is mediated by HSPGs. Heparin appears to be a  
competitor and retain proheparanase in the extracellular medium  
(By similarity).  
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form  
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa  
CC product. The active form, the 8/50 kDa heterodimer, is resistant  
CC to degradation. Complete removal of the linker peptide appears to  
CC be a prerequisite to the complete activation of the enzyme (By  
CC similarity).  
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears  
CC to be essential for its solubility (By similarity).  
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL: AF39508; AA015189.1; -; mRNA.  
CC EMBL: AF184967; AF04563.1; -; mRNA.  
CC RGD: 61969; HpsE.  
DR InterPro: IPR005199; Glyco\_hydro\_79\_N.  
DR Pfam: PF03662; Glyco\_Hydro\_79n; 1.  
KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;  
KW Signal.  
FT SIGNAL 1 28 By similarity.  
FT CHAIN 29 102 Heparanase 8 kDa subunit.  
FT PROPEP 103 150 /FTid=PRO\_0000042266.  
FT FT Linker peptide (By similarity).  
FT FT /FTid=PRO\_0000042267.  
FT CHAIN 151 536 Heparanase 50 kDa subunit.  
FT FT /FTid=PRO\_0000042268.  
FT REGION 151 155 Heparin/HS-binding (By similarity).  
FT REGION 263 273 Heparin/HS-binding (By similarity).  
FT ACT\_SITE 218 218 Proton donor (Potential).  
FT ACT\_SITE 336 336 Nucleophile (Potential).  
FT CARBOHYD 155 155 N-linked (GlcNAc...) (By similarity).  
FT CARBOHYD 193 193 N-linked (GlcNAc...) (By similarity).  
FT CARBOHYD 210 210 N-linked (GlcNAc...) (By similarity).  
FT CARBOHYD 452 452 N-linked (GlcNAc...) (By similarity).  
FT CONFLICT 15 15 G -> R (in Ref. 2).  
FT CONFLICT 227 227 H -> Q (in Ref. 2).  
FT CONFLICT 350 350 D -> N (in Ref. 2).  
SQ SEQUENCE 536 AA; 60480 MW; C434E04CF536EA4D CRC64;  
Query Match 75.3%; Score 2138; DB 1; Length 536;  
Best Local Similarity 76.3%; Pred. No. 2.3e-152;  
Matches 408; Conservative 49; Mismatches 78; Indels 0; Gaps 0;  
QY 9 LPPLMLLLGLPLGPGALPRPAQADVVLDLFFQEPHLVSPSPSLVTIDANLATD 68  
Db 2 LRPLLLMLWGRLCALTQGTTPAGTAPTAKDVVDLEFYTAKLRFQSVSPSLTIDASLATD 61  
QY 69 PRELILGSPKLTARGLSPAYLRFQGTDTFLIFDPKKESTFEERSYQSVQVNDICK 128  
Db 62 PRELFTFLGSPRLALARGSLPAYLRFQGTDTFLIFDPKKESTFEERSYQSVQVNDICK 121  
QY 129 YGSIPTDVEEKLRLWEPYQEQILLRHHYQKFKNSTYSRSSVDVLYTPANCSSGLDIFGL 188  
Db 122 SERVSADVLRLKQWENFQELLRLREQYQRFKNSTYSRSSVDVLYTPANCSSGLDIFGL 181  
QY 189 NALLRTADLQWSSNAQLLDYCSSKGVNLSWELGNPNPSFLKKADIFINGSLQGEFTIQ 248  
Db 182 NALLRTADLQWSSNAQLLDYCSSKGVNLSWELGNPNPSFLKKADIFINGSLQGEFTIQ 241  
QY 249 LHKLLRSTFKNAKLYGPDVQGPQRRKTAQMLKSLKAGGEVIDSVTWHYYLNGTATRE 308  
Db 242 LHKLLRSTFKNAKLYGPDVQGPQRRKTAQMLKSLKAGGEVIDSVTWHYYLNGTATRE 301  
QY 309 DELNPVDLIFISSVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 368  
Db 302 DFLSSDVLDTFILSVQKLYLTKEMTPGKVLGETSAYGGAPLLSDTTPAAGFWLWDLK 361  
QY 369 LGLSARMGIEVWQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 428

```

Db 362 LGLSAQGLGIEVWQVFFGAGNYHLVDENFEPLDYWLSLLFKLGVKULMSRVKGPDR 421
Qy 429 RKLRYVYLHCTNTDNPRYKGGDLTYAIALNHVNTKYLRLPVPFSSKNQVDKYLRLPLGPHGL 488
Db 422 SKRLRYVYLHCTNVVHPRYREGDLTYLVNLHNVTKLKLPPPMFSPKVDKYLRLKPFSGDGL 481
Qy 489 LSKSVQLNGLTAKMVDOTLPLMEKPLRGSSGLGPAFSPFVIRNAKVAACI 543
Db 482 LSKSVQLNGTLKMWDEQTLPALTEKPLPAGSLVSPAFSGYGFVIRNAKIAACI 536

RESULT 10
HPSE_CHICK STANDARD; PRT; 523 AA.
AC Q90YK5;
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-DEC-2001, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Heparanase precursor (EC 3.2.-.-).
GN Name=HPSE; Synonyms=HPA;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archaea; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RX MEDLINE=21369959; PubMed=11387326; DOI=10.1074/jbc.M102462200;
RA Goldshmidt O., Zcharia E., Angorn H., Guatta-Rangini Z., Atzmon R.,
RA Michal I., Fecker I., Mitran E., Vlodavsky I.;
RT "Expression pattern and secretion of human and chicken heparanase are
RT determined by their signal peptide sequence.";
RL J. Biol. Chem. 276:29178-29187(2001).
CC -!- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans (By similarity).
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted
CC (By similarity).
CC -!- PTM: N-glycosylated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR ENBL; AY037007; AAK82648.1; -; mRNA.
DR Ensembl; ENSGALG0000011203; Gallus gallus.
DR InterPro; IPR005199; Glyco_hydro_79_N.
DR Pfam; PF03662; Glyco_hydro_79n; 1.
KW Glycoprotein; Hydrolase; Lysosome; Membrane; Signal.
FT SIGNAL 1 18 Potential.
FT CHAIN 19 523 Heparanase.
FT FTID=PRO_0000042259.
FT REGION 137 141 Heparin/HS-binding (By similarity).
FT REGION 250 260 Heparin/HS-binding (By similarity).
FT ACT_SITE 204 204 Proton donor (Potential).
FT ACT_SITE 323 323 Nucleophile (Potential).
FT CARBOHYD 141 141 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 196 196 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 436 436 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 439 439 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF881 CRC64;
Query Match 58.0%; Score 1648.5; DB 1; Length 523;
Best Local Similarity 60.3%; Pred. No. 2e-115;
Matches 321; Conservative
Qy 13 LMILLGLPLGFLSPALPRAQADVDVLDFFTOEPLHLVSPSLSVTIDANLATDREL 72
Db 2 LVLLLLVLLAVP-----RRTAELQLGLREPTGAVSPAFSLTLDASLATPRFV 52
Qy 73 ILGSPKRLTLARGLSPAYLRFQGTKTDFLIFDPKKESTPEERSYQSQVNQDICKYGI 132

```

```

Db 53 ALLRHPKLTLASGLSPGLFRFGTSTDFLFNPKNKSTWEEKVLSFQA-KDVCENWPS 111
Qy 133 PPVVEEKLRLWPYQEOQLLEHNYQKPKNSTYGRSSVDVLYTFANCGLDLIFGLNALL 192
Db 112 FAVVPKLLLTQWPLQEKLLAEHSWKXKNTTITRSTDLILHTFASSSGFRLVFGLNALL 171
Qy 193 RTADLQWSSNAQLLDYCCSSKGYNISWELGNERNFLKKADIFINGSOLGDEFTQLHKL 252
Db 172 RRAGLQWSSNAQLLDYCAQRSYNISWELGNERNFRKKSGICIDGDFGLRDFVHLRQL 231
Qy 253 L-RKSTFKNAKLYGPDVGQPRRKAKMLKSLKAGGEVIDSVTWHYYLNGRTATREDFL 311
Db 232 LSQPLRYHAELVGLDVGQPRKHTQHLLRSPMKSGKAIDSVTWHYYNGRSATREDFL 291
Qy 312 NPVDLDIFISSVQVQVVESTPRGKVMWLGTSAYGGAPLLSDTFAAGFMWLDKGL 371
Db 292 SPEVLDSPATAIHDVLGIVEATVPGKVMWLGTSAYGGAPQLSNTYVAGFMWLDKGL 351
Qy 372 SARNGIEVWQVFFGAGNYHLVDENFEPLDYWLSLLFKLGVTKVLMASVQSGKRRKL 431
Db 352 AARREGIDVWVRQVSGAGSYHLVDAGFKPLPDYWLSLLFKLGVTRVLQASVEQADARRP 411
Qy 432 RYVLHCTNTDNPRYKGGDLTYAIALNHVNTKYLRLPYFPFSKNQVDKYLRLPLGPHGLSK 491
Db 412 RYVLHCTNPRPKYREGDVLFAALNSVNTQSLQPKQLWKSVDQYLLPHGKDSILSR 471
Qy 492 SVQLNGLTAKMVDOTLPLMEKPLRGSSGLGPAFSPFVIRNAKVAACI 543
Db 472 EVQLNGLLQWVDDETLPALEHMAALPGSTGLPFAFSGYGFVIRNAKIAACI 523

RESULT 11
Q4SYF6 TETNG PRELIMINARY; PRT; 533 AA.
AC Q4SYF6;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 4.
DE Chromosome undetermined SCAF12073, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG00010356001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jallou O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthonard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Bismont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.

```



[illegible]

DR	GO:0005622; C:intracellular; TAS.
DR	GO:0003035; F:heparanase activity; TAS.
DR	InterPro: IPR005199; Glyco hydro 79 N.
DR	Pfam: PF03662; Glyco_hydro_79n; T.
KW	Alternative splicing; Hydrolase; Membrane; Polymorphism.
FT	CHAIN 1 592
FT	Heparanase-2.
FT	/FTid=PRO_0000068140.
FT	ACT SITE 262 262 Proton donor (Potential).
FT	VARSPLIC 150 261 Missing (in isoform 4).
FT	/FTid=VSP 015850.
FT	VARSPLIC 204 261 Missing (in isoform 3).
FT	/FTid=VSP_015851.
FT	VARSPLIC 539 548 SVOLNQGPLV -> TQRCQVCGII (in isoform 2).
FT	/FTid=VSP 015852.
FT	VARSPLIC 549 592 Missing (in isoform 2).
FT	/FTid=VSP 015853.
FT	VARIANT 579 579 F -> Y (in dbSNP:10883100).
FT	/FTid=VAR_023601.
FT	CONFLICT 12 12 P -> L (in Ref. 2; CAC82492).
FT	CONFLICT 213 213 F -> S (in Ref. 2).
SQ	SEQUENCE 592 AA; 66580 MW; 95C384AD9A74258E CRC64;
Query Match 40.5%; Score 1150.5; DB 1; Length 592;	
Best Local Similarity 43.4%; Pred. No. 9.2e-78;	
Matches 249; Conservative 83; Mismatches 189; Indels 53; Gaps 9;	
Qy	20 PLGLPLPGAL-----PPRA-----QAQDVVDLDFFTOEPLHLVSPS 55
Dd	18 PPACLAFAGLYALLHLHLSLSSOGRRLPVDRRAAGLEKTKILLDVTSTKNPTVTNEN 77
Qy	56 FLAVTIDANLATDPRELIILGSPKLRTLAGLSPAYLRFGTKTDLFIF----DPKEST 111
Dd	78 FLSLQLDPSIHD-GWLDDFLSKRLVTLARGLSPAFLRFGGKRDTDLFQNLNRPAKSRG 136
Qy	112 FEERSYWSQVNODI-----CKYGSTPPDVEEKRLRWEPYQEOL--LLREHYOK 158
Dd	137 GPGPDYLYLNKYEDIVRSOVALEDKQKGIAQ-HPDVMLELOREKAQMHLVLLEKQFSN 195
Qy	159 KFKNSYTSRSSVDLVYTFANCGLDIIFGLNALLRTADLQWNSSNAQLLLDYCSSKGYNI 218
Dd	196 TYSNLILTARSLDKLYNFADCGLHLIFALNALRRPNNSWNSSALSLLKYTSASKYNI 255
Qy	219 SWELGNEPSFLKKADI FINGSGLEDFTQLHKLRK-STFKNAKLYGPDVGQPRRKTA 277
Dd	256 SWELGNENPYRTMHGRAVNGSLGQDIQLSKLLQPIRIYSRASLYGPNIGRPKNVIA 315
Qy	278 MLKSFUKAGEVIDSVTHWHYLNGRTATREDLPNDVLDFDISSVKVQFOVVESTRPGK 337
Dd	316 LLDGFNMKAGSTVDAVTWQHCVIDGRVWKVMDPLKTRLLDTSQDIRKI QKVNTYTTPGK 375
Qy	338 KWMLGETSAYGGAPLLSDTTAAQFMWLDKLGSLARMGLEVVMRQVFEGAGNYHLVDEN 397
Dd	376 KIWEGVVITSAGTNLSDSYANGFLMWLTGLMANQGIDVVIRHSFFDHGYNHLVDQN 435
Qy	398 FDPPLPYWLSLAFKKLVGTKLMASVQGSKRR-----KLRVYLHCITNTDNPRYKEG 448
Dd	436 FNPLDPYWLSLYKRLIGPKVLAVHAGLQRPRGPRGVRIRDKLRIYAHCINHHNNYVRG 495
Qy	449 DLTLYAINLHNVTYLRYPSPFNKOVDKYLRPLRPLGHGLLSKSVOLNGJTLKQVDDQTLL 508
Dd	496 SITLFTINLHRSKTKTAGLTRDKLVHQVLLQPYQGEGLKSKSVOLNGPLVMVDDGTL 555
Qy	509 PPLMEPLRPGSSLGLPAFSYFFEVIRNAKVAAC 542
Dd	556 PELKPRPLRAGR TLVIPPTMGFFVVKVNVALLAC 589
RESULT 13	
ID	Q2M1H9 HUMAN PRELIMINARY; PRF; 592 AA.
AC	Q2M1H9; AC
DT	21-FEB-2006, integrated into UniProtKB/TrEMBL.
DT	21-FEB-2006, sequence version 1.





Db	72	NFLSLQDPSI IKD -GWLD FSSGRLVLTARGLSPAFLRFGGRKTDFQFTNQKNLAKPR	130
Qy	110	---STFEERSYQWOSVNODI-----CKYGSIPDPDVEKRLWEYPYEQFL--LLREHYQK	158
Db	131	GPQPDYILKNYEDDIIRSDIALDKQKGCKLAS-HPDMLELQREKAASTQLVLLKEQLSN	189
Qy	159	KFKNST----YSRS-----SVDVLYTFANCSDLGLIFGLNALLRTADLOWNSSNAQLLLDY	210
Db	190	IYSNITLTGLFSHRSIARSLDKLYNFADOCAGLHLILGLNALHRNPDHSWNTSSTLSLLKY	249
Qy	211	CSSKGYNISWELGNEPNSFLKKAADIFINGSOLGEDFIOLHKLRLK-SFFKNAKLYGPDVG	269
Db	250	SAGKKYNISWELGNEPNAYSRWSVGHAVNSSLQAQDYTKLRTLQSVYYSRAQLGYPNAG	309
Qy	270	QPRBKTAOKLSFKKAGGEVIDSVTHWHYILNGRTATREDFLNPDVLDIFISSVQKVQVQ	329
Db	310	RPRKNALLLDEPFMKTGTVVDVATVQHYHMDGRIKKYVEDFLKTELDTLTLEQLSKVTKY	369
Qy	330	YESTRPGKKWLGWETSSAYGGGAPLLSTTFAAGFMWLDKLGLSARMGIEVVMR-----QVF	385
Db	370	VNHTPGKKWLGGLGPAMTGMNSLSDTFAAGFLWNTLGMAAQGIDVVLRRQAVQEH	429
Qy	386	PGAGNYHLVDENFDP-LPDYWL SLLFKLVGTKVLMASVQSGKRP-----KLRYVL	435
Db	430	TNKOSVALFLQMFVPSPDYFSLVFKELVGPKVLA VRVAGLQRKQPQGRVIRDKLRYA	489
Qy	436	HCTWTDNPRYKEGDLTLVAINLHNVTYLRLYPPSPSNQVDKYLLRPLGPHGLLSKSVQL	495
Db	490	HCTSYSNHNYVRSGITIYIINLHRSKKIKLAGTLRNVIHQYLLOPYGADGLRAKRVQL	549
Qy	496	NGLTLKMVDODTLPLMEKPLRPQSSSLGPLAFSFFVIRNAKVAAC	542
Db	550	NGEKLLMADNETFPELKPKTIRAGRTIAMPBMTTIGFYVKKINAYAC	596

## RESULT 15

Q4TGC8\_TETNG PRELIMINARY; PRT; 255 AA.  
ID Q4TGC8\_TETNG  
AC Q4TGC8;  
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
DT 19-JUL-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Chromosome undetermined SCAF3783, whole genome shotgun sequence.  
DE (Fragment).  
DE ORFNames=GSTENG00001168001;  
GN Tetraodon nigroviridis (Green puffer).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
OC Tetraodontidae; Tetraodontidae; Tetraodon.  
OX NCBI\_TaxID=99883;  
OX [1].  
RN NUCLEOTIDE SEQUENCE.  
RP PubMed=15496914; DOI=10.1038/nature03025;  
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,  
RA Mauceli E., Bouteau L., Fischer C., Ozouf-Costaz C., Bernot A.,  
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,  
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
RA Anthonard V., Jubin C., Castelli V., Katinka M., Vacherie B.,  
RA Blemont C., Skalli Z., Catolico L., Poulain J., De Berardinis V.,  
RA Craud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,  
RA Parra G., Hardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,  
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,  
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
RA Wincker P., Lander E.S., Weissbach J., Roest Croliius H.;  
RT "genome duplication in the teleost fish Tetraodon nigroviridis reveals  
RT the early vertebrate proto-karyotype.";  
RL Nature 431:946-957(2004).  
RL [2].  
RN NUCLEOTIDE SEQUENCE.  
RG Genoscope; Whitehead Institute Centre for Genome Research;

RL		Submitted (FEB-2004) to the EMBL/GenBank/DDBJ databases.					
CC	-!	CAUTION: The sequence shown here is derived from an					
CC	EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is preliminary data.						
CC		-----					
CC		Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms					
CC	Distributed under the Creative Commons Attribution-NoDerivs License						
CC		-----					
DR	EMBL; CAAE01003783; CAF88054.1.; -; Genomic_DNA.						
FT	NON TER            1						
FT	NON TER         255						
SQ	SEQUENCE      255 AA;	28562 MW;	07F542A9C755E3F0 CRC64;				
		Query Match          26.1%;	Score 742.5;	DB 2;	Length 255;		
		Best Local Similarity 56.7%;	Pred. No. 1.6e-47;				
		Matches 144;	Conservative 37;	Mismatches 56;	Indels 17; Gaps 3		
Qy	166	SRSSVDVLYTFANGCSGLDLIFGLNALLRATDQLWNSSNAQLLLDYCSSKGYNISWEELGNE	225				
Db	1	SETTVDQLHAFANCSGLDLLFVLGNALLRTADNRNSSNARSILRRYCEARRHYMSHELNGE	60				
		: : : : : :	:	:	:		
Qy	226	PNSEFLKKADIFINGSOLGEDEPIQLHKLLRKSTP-KNAKLYGDVGQQPRRKTANKMLSKFLK	284				
Db	61	PNSEYEKAGLRDLGRQLGEDFTVLRIKRSLRESFYRDAGLFPGPDVGQPRDHRIIDLSGFLO	120				
		: : : : : :	:	:	:		
Qy	285	AGEVIDSVTHWHYYLNGRTRATRDEFLNPVDLDIIFISSVKVFQVVESTRPCKKYWLGET	344				
Db	121	SGAEAVDACHTWHYYLDGREASLEDFDDPDVDLTTLREKIGEVLEEHVQSPGPKYWLGET	180				
		: : : : : :	:	:	:		
Qy	345	SSAYGGG-----APLLSDTTFAAG-FMWLLDKLGLSARMGIEVVMRVQFFGA	388				
Db	181	SSATGAEPRCRTHSSQDSCEFAFRSDDQAPLGTFRWLDTKLGLAATLGLVMRQVLIGA	240				
Qy	389	GNYHLVDENFDPLP	402				
Db	241	GSYHLMDDNLDP LP	254				
		: : : : : :	:	:	:		

Search completed: June 5, 2006, 12:20:12  
Job time : 137.728 secs



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:06 ; Search time 106.86 Seconds  
(without alignments)  
2293.354 Million cell updates/sec

Title: US-10-645-659A-3  
Perfect score: 2800  
Sequence: 1 MLRPLLLWGLRLALTGQ.....VPAFSYGFVIRNAKIRACI 536

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_8:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003s:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*
- 9: Geneseqp2005s:\*
- 10: Geneseqp2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2800	100.0	536	5 ABB07812	Abb07812 Rat hepar
2	2800	100.0	536	8 ADR88209	Adr88209 Rat hepar
3	2800	100.0	536	8 ADT78176	Adt78176 Rat hepar
4	2800	100.0	536	9 ADY27035	Ady27035 Rat hepar
5	2800	100.0	536	9 AEA42425	Aea42425 Rat hepar
6	2594.5	92.7	535	9 ADY27033	Ady27033 Murine he
7	2590.5	92.5	535	3 Aab08851	Aab08851 A murine
8	2590.5	92.5	535	3 ABB07811	Abb07811 Mouse hep
9	2590.5	92.5	535	7 ADG88834	Adg88834 Mouse hpa
10	2590.5	92.5	535	8 ADL16413	Adl16413 Mouse hep
11	2590.5	92.5	535	8 ADM48750	Adm48750 Mouse hpa
12	2590.5	92.5	535	8 ADR88208	Adr88208 Mouse hep
13	2590.5	92.5	535	8 ADT78175	Adt78175 Mouse hep
14	2590.5	92.5	535	9 AEA42424	Aea42424 Mouse hep
15	2151	76.8	543	2 ADY27034	Ady27034 Bovine hep
16	2126	75.9	543	2 AAY17082	Aay17082 Human hep
17	2126	75.9	543	4 AAB86206	Aab86206 Human hep
18	2126	75.9	543	7 ADD18950	Adl18950 Human dis
19	2126	75.9	543	8 ADK52086	Adk52086 Human ato
20	2126	75.9	543	8 ADM48759	Adm48759 Human hpa
21	2126	75.9	543	8 ADN05074	Adn05074 Antipsori
22	2126	75.9	543	8 ADN04902	Adn04902 Antipsori
23	2126	75.9	543	8 ADQ80372	Adq80372 Heparanas

24	2126	75.9	543	8 ADR88210	Adr88210 Human pre
25	2126	75.9	543	8 ADP25079	Adp25079 PRO poly
26	2126	75.9	543	8 ADT78177	Adt78177 Human hep
27	2126	75.9	543	9 ADY27036	Ady27036 Human hep
28	2126	75.9	543	9 AEA42426	Aea42426 Human hep
29	2126	75.9	588	2 AAY30124	Aay30124 A human p
30	2125.5	75.9	545	6 ABP56822	Abp56822 Human hep
31	2125.5	75.9	545	7 ADE16012	Ade16012 G-coupled
32	2125.5	75.9	545	8 ADL93951	Adl93951 Human G-c
33	2123	75.8	543	2 AAY02345	Aay02345 A human h
34	2123	75.8	543	3 AAY57590	Aay57590 Human hep
35	2123	75.8	543	3 AAB08849	Aab08849 Amino aci
36	2123	75.8	543	3 AAY52990	Aay52990 Human hep
37	2123	75.8	543	4 AAY97635	Aay97635 Human hep
38	2123	75.8	543	5 ABB07813	Abb07813 Human hep
39	2123	75.8	543	7 ADG88800	Adg88800 Human hpa
40	2123	75.8	543	8 ADL16379	Adl16379 Human hep
41	2123	75.8	543	8 ADM48716	Adm48716 Human hpa
42	2123	75.8	543	9 AEA42466	Aea42466 Human hep
43	2123	75.8	543	10 AEE96848	Aee96848 Human hep
44	2123	75.8	592	2 AAY02346	Aay02346 A human h
45	2123	75.8	592	3 AAB08850	Aab08850 Amino aci

#### ALIGNMENTS

RESULT 1	ABB07812	ID	ABB07812 standard; protein; 536 AA.
XX	AC	ABB07812;	
XX	DT	03-JUL-2002 (first entry)	
XX	DE	Rat heparanase sequence.	
XX	XX	Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme; anti-protozoan; neuroprotective; heparin; rat.	
XX	OS	Rattus rattus.	
XX	XX	Key	Location/Qualifiers
FT	Peptide	1..16	
FT	Protein	/note= "putative signal peptide"	
FT		17..536	
FT		/note= "mature protein"	
XX	FN	US2002034810-A1.	
XX	PD	21-MAR-2002.	
XX	PF	16-AUG-2001; 2001US-00930218.	
XX	PR	20-SEP-2000; 2000US-00666390.	
XX	PA	(INSI-) INSIGHT STRATEGY & MARKETING LTD.	
XX	PI	Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zeharia E;	
XX	DR	WPI; 2002-338926/37.	
XX	PT	Nucleic acid encoding avian and reptile heparanase polypeptide is useful to treat various heparin-related disorders and the signal peptide is useful in production of membrane-targeted or secreted recombinant proteins.	
XX	PS	Disclosure; Fig 1a; 39pp; English.	
XX	CC	The invention relates to an isolated avian and reptile nucleic acid, encoding a polypeptide with heparanase catalytic activity. The signal peptide of the nucleic acid can be used to express membrane-associated or secreted proteins in heterologous expression systems. The encoded	

CC polypeptides can be used to prevent tumour angiogenesis, metastasis and  
CC invasion, and to intervene with pathologies associated with impaired  
CC heparin-binding growth factors, cellular responses to heparin-binding  
CC growth factors and cytokines, cell interaction with plasma lipoproteins,  
CC cellular susceptibility to viral, protozoa and bacterial infections or  
CC disintegration of neurodegenerative plaques. The present sequence  
CC represents a rat heparanase protein sequence used in similarity studies  
XX  
SQ Sequence 536 AA;

Query Match 100.0%; Score 2800; DB 5; Length 536;  
Best Local Similarity 100.0%; Pred. No. 8.6e-260; Indels 0; Gaps 0;  
Matches 536; Conservative 0; Mismatches 0;

QY 1 MLRPLLLMLWGLRALTGCTPAGTAKDQVLEFFYTKELFQSVSPFLSIITDASLAT 60  
DB 1 MLRPLLLMLWGLRALTGCTPAGTAKDQVLEFFYTKELFQSVSPFLSIITDASLAT 60

QY 61 DPRFLTFGLSPRLRALARGLSPAYLRGGTKTDFLIPDPNKEPTSEERSYQSDNDIC 120  
DB 61 DPRFLTFGLSPRLRALARGLSPAYLRGGTKTDFLIPDPNKEPTSEERSYQSDNDIC 120

QY 121 GSERVSADVLKQWEPFQELLRLREYQREFKNSTYSRSSVDMLYSPAKCSRLDLIFG 180  
DB 121 GSERVSADVLKQWEPFQELLRLREYQREFKNSTYSRSSVDMLYSPAKCSRLDLIFG 180

QY 181 LNALLTPDLRWNSNAQLLNTCSSGNYISWELGNEPNSFWKKAQISIDGLQLGDFV 240  
DB 181 LNALLTPDLRWNSNAQLLNTCSSGNYISWELGNEPNSFWKKAQISIDGLQLGDFV 240

QY 241 ELHKLQKSAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGGEVDSLTWHHYLNGRVATK 300  
DB 241 ELHKLQKSAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGGEVDSLTWHHYLNGRVATK 300

QY 301 EPLSSDVLDTFLSVQKILKVTKEPMPGKVMWLGTSYAGGAPLLSNTFAAGFWMLD 360  
DB 301 EPLSSDVLDTFLSVQKILKVTKEPMPGKVMWLGTSYAGGAPLLSNTFAAGFWMLD 360

QY 361 KLGLSQGLGVVMRVQVFFGAGNYHLDENFEPPLDYWLSLLPKLGGPKVLMRSVKGPD 420  
DB 361 KLGLSQGLGVVMRVQVFFGAGNYHLDENFEPPLDYWLSLLPKLGGPKVLMRSVKGPD 420

QY 421 RSKLRVYLHCTNVYHPRYREGDITLVYVNLHNTVTKHLKLPFPMFSRVDVKYLLKPFSGDG 480  
DB 421 RSKLRVYLHCTNVYHPRYREGDITLVYVNLHNTVTKHLKLPFPMFSRVDVKYLLKPFSGDG 480

QY 481 LLSKSVOLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPAPSYGFFVIRNAKIAACI 536  
DB 481 LLSKSVOLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPAPSYGFFVIRNAKIAACI 536

RESULT 2  
ADR88209  
ID ADR88209 standard; protein; 536 AA.  
XX  
AC ADR88209;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Rat heparanase.  
XX  
KW Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;  
KW autoimmune disorder; cancer; angiogenesis; metastatic disease;  
KW atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;  
KW haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;  
KW Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;  
KW rat; heparanase; enzyme.  
XX  
OS Rattus norvegicus.  
XX  
FH  
FT Peptide  
FT 1. .16  
FT /label= Signal\_peptide

FT Protein 17. .536  
FT /label= Mature\_heparanase  
XX US2004170631-A1.  
XX 02-SEP-2004.  
XX 28-NOV-2003; 2003US-00722502.  
XX 02-SEP-1997; 97US-00922170.  
XX 01-MAY-1998; 98US-00071739.  
XX 04-NOV-1998; 98US-00186200.  
XX 19-FEB-2003; 2003US-00368044.  
XX 22-AUG-2003; 2003US-00645659.  
XX (YACO/) YACOBY-ZEEVI O.  
PA (PERE/) PERETZ T.  
PA (MIRO/) MIRON D.  
PA (SHLO/) SHLOMI Y.  
PA (PECK/) PECKER I.  
PA (AYAL/) AYAL-HERSHKOVITZ M.  
PA (FEIN/) FEINSTEIN E.  
PA (VCEL/) VAN GELDER J M.  
PA (VLOD/) VLODAVSKY I.  
PA (FRIE/) FRIEDMANN Y.  
XX  
PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;  
PI Friedmann Y;  
XX WPI; 2004-625084/60.  
XX  
XX Targeted drug delivery to a heparanase-expressing tissue of a patient,  
XX useful for treating heparanase-associated conditions such as inflammation  
XX or cancer, comprises administering a drug and an anti-heparanase antibody  
XX complex.  
XX  
XX Claim 2; SEQ ID NO 3; 58pp; English.  
XX  
XX The invention relates to a method of targeted drug delivery to a tissue  
XX of a patient, the tissue expressing heparanase. The method comprises  
XX providing a complex of a drug directly or indirectly linked to an anti-  
XX heparanase antibody, and administering the complex to the patient. In the  
XX targeted drug delivery, the antibody comprises an antibody or its portion  
XX capable of specifically binding to at least one epitope of a heparanase  
XX protein. The composition and methods of the invention are useful for  
XX diagnosing, preventing or treating conditions associated with heparanase  
XX catalytic activity (e.g. an inflammatory disorder, wound, scar,  
XX vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell  
XX proliferation, invasion of circulating tumour cells and metastatic  
XX disease) for purifying heparanase, or for developing drugs for those  
XX heparanase-associated conditions. The vasculopathy is atherosclerosis,  
XX restenosis or aneurysm. The cancerous condition is a solid cancer or a  
XX non-solid cancer. The non-solid cancer is a haematopoietic malignancy  
XX selected from acute lymphocytic leukaemia (ALL), acute myelogenous  
XX leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous  
XX leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,  
XX Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and  
XX multiple myeloma. The solid cancer is selected from tumours in lip and  
XX oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,  
XX thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,  
XX anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of  
XX Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue  
XX sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,  
XX vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic  
XX tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary  
XX bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,  
XX malignant melanoma of the conjunctiva, malignant melanoma of the uvea,  
XX retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,  
XX brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's  
XX sarcoma. The present sequence is rat heparanase.  
XX  
XX Sequence 536 AA;

Query Match	100.0%;	Score 2800;	DB 8;	Length 536;
Best Local Similarity	100.0%;	Pred. No. 8.6e-260;	Indels 0;	Gaps 0;
Matches 536;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MLRPLLLWLGRLALQTGTPAGTAPTKDVVDLEFYTKRLFQSVSPSLITIDASLAT	60	
DB	1	MLRPLLLWLGRLALQTGTPAGTAPTKDVVDLEFYTKRLFQSVSPSLITIDASLAT	60	
QY	61	DPRLFTFLGSPRLALARGLSPAYLRFGGTKTDFLIIDPNKPTSEERSYQSQNDNDIC	120	
DB	61	DPRLFTFLGSPRLALARGLSPAYLRFGGTKTDFLIIDPNKPTSEERSYQSQNDNDIC	120	
QY	121	GSRVSADVLRLQWEPFQELLRLREYQREFKNSTYSRSSVDMLYSFAKCSRLDLIFG	180	
DB	121	GSRVSADVLRLQWEPFQELLRLREYQREFKNSTYSRSSVDMLYSFAKCSRLDLIFG	180	
QY	181	LNALLRTPDLRWNSSNAQLLNLCSSKGYNISWELGNEPNSFWKKAQISIDGLQLGDFV	240	
DB	181	LNALLRTPDLRWNSSNAQLLNLCSSKGYNISWELGNEPNSFWKKAQISIDGLQLGDFV	240	
QY	241	ELHKLQKSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYLLNGRVATK	300	
DB	241	ELHKLQKSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYLLNGRVATK	300	
QY	301	EDPLSSDVLDTFLSVQKILKVTKEMTPGKVMWLGTSYAGGAPLLSNTFAAGFWMLD	360	
DB	301	EDPLSSDVLDTFLSVQKILKVTKEMTPGKVMWLGTSYAGGAPLLSNTFAAGFWMLD	360	
QY	361	KGLSLAQGLGVVMRQVFFGAGNHYLVNDFEPLPDYWLISLLFKLGPVLMRSVRKGP	420	
DB	361	KGLSLAQGLGVVMRQVFFGAGNHYLVNDFEPLPDYWLISLLFKLGPVLMRSVRKGP	420	
QY	421	RSKLRYLHCTNVYHPYREGDLTYVLNLHNTVKHLKLPMPFSRPPVDKYLKPFQSDG	480	
DB	421	RSKLRYLHCTNVYHPYREGDLTYVLNLHNTVKHLKLPMPFSRPPVDKYLKPFQSDG	480	
QY	481	LLSKSVQLNGQTLKWVDEQTLPALTEKPLPAGSSLSVPAFSYGFFVIRNAKIAACI	536	
DB	481	LLSKSVQLNGQTLKWVDEQTLPALTEKPLPAGSSLSVPAFSYGFFVIRNAKIAACI	536	
RESULT 3				
ADT78176				
ID	ADT78176	standard; protein; 536 AA.		
XX				
AC	ADT78176;			
XX				
DT	13-JAN-2005	(first entry)		
XX				
DE	Rat heparanase protein.			
XX				
KW	Antibody; epitope; heparanase; pathological condition; angiogenesis;			
KW	cell proliferation; cancerous condition; tumour cell invasion;			
KW	metastatic disease; heparanase-related disorder; inflammatory disorder;			
KW	wound; scar; vasculopathy; autoimmune condition; renal disease;			
KW	cytostatic; antiinflammatory; vulnerrary; antiarteriosclerotic;			
KW	vasotropic; immunosuppressive; nephrotropic; antidiabetic; rat.			
OS	Rattus norvegicus.			
XX				
FT	Key	Location/Qualifiers		
FT	Binding-site	150..155		
FT	Binding-site	/note= "Putative heparin binding site"		
FT	Binding-site	264..270		
FT	Binding-site	/note= "Putative heparin binding site"		
FT	Binding-site	419..426		
FT	Binding-site	/note= "Putative heparin binding site"		
XX				
PN	US2004213789-A1.			
XX				
PD	28-OCT-2004.			
XX				

PF	22-AUG-2003;	2003US-00645659.		
XX				
FR	02-SEP-1997;	97US-00922170.		
PR	01-MAY-1998;	98US-00071739.		
PR	04-NOV-1998;	98US-00186200.		
PR	19-FEB-2003;	2003US-00368044.		
XX				
PA	(YACO/) YACOBY-ZEEVI O.			
PA	(PERE/) PERETZ T.			
PA	(MIRO/) MIRON D.			
PA	(SHLO/) SHLOMI Y.			
PA	(PECK/) PECKER I.			
PA	(AYAL/) AYAL-HERSHKOVITZ M.			
PA	(FEIN/) FEINSTEIN E.			
PA	(GELD/) GELDER J M V.			
PA	(VLOD/) VLODAVSKY I.			
PA	(FRIE/) FRIEDMANN Y.			
XX				
PI	Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;			
PI	Ayal-Herschkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;			
PI	Friedmann Y;			
DR	WPI; 2004-774790/76.			
XX				
PT	New neutralizing monoclonal anti-heparanase antibodies, useful for			
PT	detecting, treating or preventing cancer, inflammatory or autoimmune			
PT	disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.			
XX				
PS	Claim 5; SEQ ID NO 3; 68pp; English.			
XX				
CC	The invention relates to an isolated antibody or antibody portion capable			
CC	of specifically binding to or elicited by at least one epitope of a			
CC	heparanase protein, where the heparanase protein is at least 60%			
CC	homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and			
CC	where at least one epitope comprises a sequence at least 70% homologous			
CC	to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)			
CC	a hybridoma cell line comprising a cell line for producing the monoclonal			
CC	antibody, (b) a method for detecting, treating or preventing a			
CC	pathological condition or a heparanase-related disorder or condition in a			
CC	subject, (c) a method for monitoring the state of a heparanase-related			
CC	disorder or condition in a subject, and (d) a pharmaceutical composition			
CC	comprising the isolated anti-heparanase antibody or antibody portion and			
CC	a pharmaceutical carrier. The antibody, methods, and composition are			
CC	useful for detecting, treating, preventing or monitoring a pathological			
CC	condition, e.g. angiogenesis, cell proliferation, a cancerous condition			
CC	(blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,			
CC	or prostate cancer), minor cell proliferation, invasion of circulating			
CC	tumour cells, or a metastatic disease, or a heparanase-related disorder			
CC	or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy			
CC	(atherosclerosis, restenosis, or aneurysm), autoimmune condition, or			
CC	renal disease or disorder (diabetic nephropathy, glomerulosclerosis,			
CC	nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell			
CC	carcinoma) in a mammal. This sequence represents rat heparanase.			
XX				
SQ	Sequence 536 AA;			
	Query Match	100.0%;	Score 2800;	DB 8; Length 536;
	Best Local Similarity	100.0%;	Pred. No. 8.6e-260;	
	Matches 536;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	1	MLRPLLLWLGRLALQTGTPAGTAPTKDVVDLEFYTKRLFQSVSPSLITIDASLAT	60	
Db	1	MLRPLLLWLGRLALQTGTPAGTAPTKDVVDLEFYTKRLFQSVSPSLITIDASLAT	60	
Qy	61	DPRLFTFLGSPRLALARGLSPAYLRFGGTKTDFLIIDPNKPTSEERSYQSQNDNDIC	120	
Db	61	DPRLFTFLGSPRLALARGLSPAYLRFGGTKTDFLIIDPNKPTSEERSYQSQNDNDIC	120	
Qy	121	GSRVSADVLRLQWEPFQELLRLREYQREFKNSTYSRSSVDMLYSFAKCSRLDLIFG	180	
Db	121	GSRVSADVLRLQWEPFQELLRLREYQREFKNSTYSRSSVDMLYSFAKCSRLDLIFG	180	
Qy	181	LNALLRTPDLRWNSSNAQLLNLCSSKGYNISWELGNEPNSFWKKAQISIDGLQLGDFV	240	

Db 181 LNALLRTPDLRWNSSNAQLLNLYCSSKGYNISWELGNEPNSFWKKAQISIDGLQLGEDFV 240  
Qy 241 ELHKLQKSAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300  
Db 241 ELHKLQKSAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300  
Qy 301 EDFLSSDVLDTFILSVQKILKVTKEPTGKVKWLGETSSAYGGAPLLSNTFAAGFWMLD 360  
Db 301 EDFLSSDVLDTFILSVQKILKVTKEPTGKVKWLGETSSAYGGAPLLSNTFAAGFWMLD 360  
Qy 361 KLGLSAQLGIEVVMRQVFFGAGNHYLVNDFEPLPDYWLISLLPKLVGPKVMSRVKGP 420  
Db 361 KLGLSAQLGIEVVMRQVFFGAGNHYLVNDFEPLPDYWLISLLPKLVGPKVMSRVKGP 420  
Qy 421 RSKLRVYLHCTNVYHPRYREGDLTYVNLNHNVTKHLKLPMPFSPVDKYLLKPFSGDG 480  
Db 421 RSKLRVYLHCTNVYHPRYREGDLTYVNLNHNVTKHLKLPMPFSPVDKYLLKPFSGDG 480  
Qy 481 LLSKSVQLNGQTLKMWDEQTLPALETEKPLPAGSSLSVPAFSYGFVIRNAKIAACI 536  
Db 481 LLSKSVQLNGQTLKMWDEQTLPALETEKPLPAGSSLSVPAFSYGFVIRNAKIAACI 536  
RESULT 4  
ID ADY27035  
AC ADY27035  
AC ADY27035;  
DT 05-MAY-2005 (first entry)  
DE Rat heparanase protein.  
KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
KW neurological disease; viral infection; infection; cytostatic;  
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
KW protease; enzyme; enzyme purification.  
OS Rattus norvegicus.  
XX WO2005016227-A2.  
XX 24-FEB-2005.  
XX 12-AUG-2004; 2004WO-IL000744.  
XX 14-AUG-2003; 2003US-0494800P.  
XX 12-JAN-2004; 2004US-0535492P.  
XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.  
XX Van-Gelder JM, Miron D;  
PI WPI; 2005-182203/19.  
XX  
PT Regulating heparanase activity, useful for treating heparanase-associated  
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
PT neurological diseases or viral diseases) comprises modulating heparanase  
PT activation.  
XX  
XX Disclosure; SEQ ID NO 7; 211pp; English.  
XX  
CC The invention relates to a method of regulating heparanase activity in a  
CC tissue or regulating a biological process depending at least in part on  
CC heparanase activity comprising modulating heparanase activation. The  
CC invention also relates to methods of treating a heparanase- or heparin  
CC binding protein-associated disease or disorder in a subject, a  
CC pharmaceutical composition for use in the treatment of a heparanase-  
CC associated disease or disorder comprising a therapeutic amount of an  
CC agent capable of modulating heparanase activation and a pharmaceutical  
CC carrier or diluent, a method of identifying a protease activator of  
CC heparanase, a protease substrate mimetic comprising a peptide

CC representing a subset or all substrate residues or cleavage sites of  
CC human heparanase or an equivalent non-human heparanase, a method of  
CC producing active heparanase and a method of modulating an adhesion  
CC activity of heparanase. The composition and methods are useful for  
CC modulating heparanase activation and for treating heparanase-associated  
CC diseases or disorders such as cancer, inflammation, cardiovascular  
CC diseases, neurological diseases or viral infections. This sequence  
CC represents a rat heparanase protein used in the scope of the invention.  
XX  
Qy Sequence 536 AA;  
Query Match 100.0%; Score 2800; DB 9; Length 536;  
Best Local Similarity 100.0%; Pred. No. 8.6e-260;  
Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MLRPLLLLLWGRRLALTQGTPTAGTAPKDVVDLEFYTKRLFQSVSPSLTIDASLAT 60  
Db 1 MLRPLLLLLWGRRLALTQGTPTAGTAPKDVVDLEFYTKRLFQSVSPSLTIDASLAT 60  
Qy 61 DPRPLTEFLGSPRLRALARGLSPAYLRFEGTKTDFLIIDPNKEPTSEERSYQSQDNNDIC 120  
Db 61 DPRPLTEFLGSPRLRALARGLSPAYLRFEGTKTDFLIIDPNKEPTSEERSYQSQDNNDIC 120  
Qy 121 GSERVSADVLRLQMEWPFQELLLLREYQREFKNSTYSRSSVDMLYSFAKCSRLDLIFG 180  
Db 121 GSERVSADVLRLQMEWPFQELLLLREYQREFKNSTYSRSSVDMLYSFAKCSRLDLIFG 180  
Qy 181 LNALLRTPDLRWNSSNAQLLNLYCSSKGYNISWELGNEPNSFWKKAQISIDGLQLGEDFV 240  
Db 181 LNALLRTPDLRWNSSNAQLLNLYCSSKGYNISWELGNEPNSFWKKAQISIDGLQLGEDFV 240  
Qy 241 ELHKLQKSAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300  
Db 241 ELHKLQKSAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300  
Qy 301 EDFLSSDVLDTFILSVQKILKVTKEPTGKVKWLGETSSAYGGAPLLSNTFAAGFWMLD 360  
Db 301 EDFLSSDVLDTFILSVQKILKVTKEPTGKVKWLGETSSAYGGAPLLSNTFAAGFWMLD 360  
Qy 361 KLGLSAQLGIEVVMRQVFFGAGNHYLVNDFEPLPDYWLISLLPKLVGPKVMSRVKGP 420  
Db 361 KLGLSAQLGIEVVMRQVFFGAGNHYLVNDFEPLPDYWLISLLPKLVGPKVMSRVKGP 420  
Qy 421 RSKLRVYLHCTNVYHPRYREGDLTYVNLNHNVTKHLKLPMPFSPVDKYLLKPFSGDG 480  
Db 421 RSKLRVYLHCTNVYHPRYREGDLTYVNLNHNVTKHLKLPMPFSPVDKYLLKPFSGDG 480  
Qy 481 LLSKSVQLNGQTLKMWDEQTLPALETEKPLPAGSSLSVPAFSYGFVIRNAKIAACI 536  
Db 481 LLSKSVQLNGQTLKMWDEQTLPALETEKPLPAGSSLSVPAFSYGFVIRNAKIAACI 536  
RESULT 5  
AEA42425  
ID AEA42425 standard; protein; 536 AA.  
XX  
XX AEA42425;  
AC AEA42425;  
XX  
DT 28-JUL-2005 (first entry)  
XX  
XX Rat heparanase epitope peptide SEQ ID NO:3.  
XX  
XX antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;  
XX antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;  
XX inflammation; wound healing; scarring; vasculopathy; autoimmune disease;  
XX angiogenesis disorder; cancer; tumor; metastasis.  
OS Rattus norvegicus.  
XX  
XX AU2004201462-A1.  
XX  
XX 06-MAY-2004.  
XX

PF 08-APR-2004; 2004AU-00201462.  
PR 08-APR-2004; 2004AU-00201462.  
PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.  
PI (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.  
XX Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;  
PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;  
PI Feinman E;  
XX WPI; 2005-173343/19.  
XX  
XX Novel isolated antibody capable of specifically binding to epitope of  
PT heparanase protein, useful for preventing and treating heparanase-related  
PT disorder such as inflammatory disorder, scars, autoimmune conditions or  
PT angiogenesis.  
XX  
XX Claim 2; SEQ ID NO 3; 260pp; English.  
XX  
XX The invention relates to an isolated antibody or its portion (I) capable  
CC of specifically binding to an epitope of a heparanase protein. Also  
CC described: (1) a cell line (II) for producing a monoclonal antibody or  
CC its portion, comprising a cell line for producing (I); (2) a  
CC pharmaceutical composition comprising (I) and a carrier; and (3) an  
CC affinity medium (III) for binding human heparanase polypeptides,  
CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)  
CC useful for treating a subject suffering from a pathological condition,  
CC which involves administering (I) to the subject. (I) is useful for  
CC preventing and treating heparanase-related disorder or condition chosen  
CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune  
CC condition, angiogenesis, cell proliferation, cancerous condition, tumor  
CC cell proliferation, invasion of circulating tumor cells and metastatic  
CC disease. (I) is useful for detecting the presence of heparanase  
CC polypeptide in a sample. (I) is useful for detecting heparanase-related  
CC disease or condition in a subject such as vertebrate; preferably mammal  
CC e.g., human. The heparanase-related disorder or condition further  
CC includes renal disease or disorder chosen from diabetic nephropathy,  
CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome  
CC and renal cell carcinoma. The present sequence represents rat heparanase,  
CC which is used in the exemplification of the present invention.  
XX  
SQ Sequence 536 AA;

Query Match 100.0%; Score 2800; DB 9; Length 536;  
Best Local Similarity 100.0%; Pred. No. 8.6e-260;  
Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRPLLLMLWGRRLALTOGTAGTAPTKDVVDLEFYTKRLFQSVSPFLSITIDASLAT 60  
DB 1 MLRPLLLMLWGRRLALTOGTAGTAPTKDVVDLEFYTKRLFQSVSPFLSITIDASLAT 60  
QY 61 DRPLFLPLGSPRLALARGLSPAYLRGGTKTDFLIIDPNKEPTSEERSYQSDNDIC 120  
DB 61 DRPLFLPLGSPRLALARGLSPAYLRGGTKTDFLIIDPNKEPTSEERSYQSDNDIC 120  
QY 121 GSERVSADVLRLQMEWPFQELLILLREYQREFKNYSRSSVDMLYSPAKCSRLDLIFG 180  
DB 121 GSERVSADVLRLQMEWPFQELLILLREYQREFKNYSRSSVDMLYSPAKCSRLDLIFG 180  
QY 181 LNALLRTPDLRWNSNAQLLLNYCCKGYNISWELGNEPNSFWKKAQIISIDGLQGEDFV 240  
DB 181 LNALLRTPDLRWNSNAQLLLNYCCKGYNISWELGNEPNSFWKKAQIISIDGLQGEDFV 240  
QY 241 ELHKLLQKSAFONAKLQPGDIPQGRGTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300  
DB 241 ELHKLLQKSAFONAKLQPGDIPQGRGTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300  
QY 301 EDFLSSDVLDTFILSVQKILKVTKEMTPGKKVWLGETSSAYGGGAPLLSNTFAAGFWMLD 360  
DB 301 EDFLSSDVLDTFILSVQKILKVTKEMTPGKKVWLGETSSAYGGGAPLLSNTFAAGFWMLD 360  
QY 361 KLGLSAQLGIEVVMRQVFFGAGNYHLVDENFEPLDYWLSLLFKLVGPKVMSRVKGPD 420

DB 361 KLGLSAQLGIEVVMRQVFFGAGNYHLVDENFEPLDYWLSLLFKLVGPKVMSRVKGPD 420  
QY 421 RSKLRVYLHCTNVVTHPRYREGDLTLVYVNLHNVTKHLKLPMPMFSRPVDKYLKPGSGDG 480  
DB 421 RSKLRVYLHCTNVVTHPRYREGDLTLVYVNLHNVTKHLKLPMPMFSRPVDKYLKPGSGDG 480  
QY 481 LLSKSVQLNGOTLKWVDSOTLPALTEKPLPAGSSLSVPAFSYGFVIRAKTAACI 536  
DB 481 LLSKSVQLNGOTLKWVDSOTLPALTEKPLPAGSSLSVPAFSYGFVIRAKTAACI 536  
RESULT 6  
ADY27033  
ID ADY27033 standard; protein; 535 AA.  
XX  
AC ADY27033;  
XX  
DT 05-MAY-2005 (first entry)  
XX  
DE Murine heparanase protein.  
XX  
KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
KW neurological disease; viral infection; infection; cytostatic;  
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
KW protease; enzyme; enzyme purification.  
XX  
OS Mus musculus.  
XX  
PN WO2005016227-A2.  
XX  
PD 24-FEB-2005.  
XX  
PF 12-AUG-2004; 2004WO-IL000744.  
XX  
PR 14-AUG-2003; 2003US-0494800P.  
PR 12-JAN-2004; 2004US-0535492P.  
XX  
PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.  
PI Van-Gelder JM, Miron D;  
XX  
DR WPI; 2005-182203/19.  
XX  
PT Regulating heparanase activity, useful for treating heparanase-associated  
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
PT neurological diseases or viral diseases) comprises modulating heparanase  
PT activation.  
XX  
PS Disclosure; SEQ ID NO 5; 211pp; English.  
XX  
CC The invention relates to a method of regulating heparanase activity in a  
CC tissue or regulating a biological process depending at least in part on  
CC heparanase activity comprising modulating heparanase activation. The  
CC invention also relates to methods of treating a heparanase- or heparin  
CC binding protein-associated disease or disorder in a subject, a  
CC pharmaceutical composition for use in the treatment of a heparanase-  
CC associated disease or disorder comprising a therapeutic amount of an  
CC agent capable of modulating heparanase activation and a pharmaceutical  
CC carrier or diluent, a method of identifying a protease activator of  
CC heparanase, a protease substrate mimetic comprising a peptide  
CC representing a subset or all substrate residues or cleavage sites of  
CC human heparanase or an equivalent non-human heparanase, a method of  
CC producing active heparanase and a method of modulating an adhesion  
CC activity of heparanase. The composition and methods are useful for  
CC modulating heparanase activation and for treating heparanase-associated  
CC diseases or disorders such as cancer, inflammation, cardiovascular  
CC diseases, neurological diseases or viral infections. This sequence  
CC represents a murine heparanase protein used in the scope of the  
XX invention.  
SQ Sequence 535 AA;

```
Query Match      92.7%; Score 2594.5; DB 9; Length 535;
Best Local Similarity 92.7%; Pred. No. 5.1e-240;
Matches 497; Conservative 19; Mismatches 19; Indels 1; Gaps 1;

QY 1 MLRPLLLWLRALRALTQGTAGTAPTDVVDLEFYTKEFLFQSVSPSFLSITIDASLAT 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MLR-LLLLLWLPGLAQAQAGTAPTDVVDLEFYTKEPLRSVSPSFLSITIDASLAT 59

QY 61 DRPFLFLGSPRLRALARGSPAYLRFGGTYKTDPLIPDPNKEPTSEERSYQSDNDIC 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 DRPFLFLGSPRLRALARGSPAYLRFGGTYKTDPLIPDPNKEPTSEERSYQSDNDIC 119

QY 121 GSRVSADVLRLQMEWPFQELLRLREYQREKFNSTYSRSSVDMLYSFACSKSLDLIFG 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 RSEPVSAVLRLQVEMPFQELLRLREYQREKFNSTYSRSSVDMLYSFACSKSLDLIFG 179

QY 181 LNALRTPLDRWNSSNAQLLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 LNALRTPLDRWNSSNAQLLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDFV 239

QY 241 ELHKLQSAFONAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 ELHKLQSAFONAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRIATK 299

QY 301 EDFLSSDVLDTFILSVQIKLVTKEMTPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 EDFLSSDVLDTFILSVQIKLVTKEMTPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 359

QY 361 KLGLSAQLGIEVVMRQVFFGAGNYHLVDENPEPLPDYWLSSLKPKLGVKPLMSRVKGP 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 KLGLSAQMGIEVVMRQVFFGAGNYHLVDENPEPLPDYWLSSLKPKLGVKPLMSRVKGP 419

QY 421 RSKLRVYLHCTNVYHPRYREGDLTYVLNLHNVTXHLKLPMPFSRPVDKYLKPFSGDG 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 RSKLRVYLHCTNVYHPRYQEGDLTYVLNLHNVTXHLKVPFPLFRKPDVYLLKPSGPG 479

QY 481 LLSKSVOLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPFASGYGFFVIRNAKIAACI 536
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 LLSKSVOLNGQILKMWDEQTLPALTEKPLPAGSALSVPFASGYGFFVIRNAKIAACI 535

RESULT 7
AAB08851
ID AAB08851 standard; protein; 535 AA.
AC AAB08851;
XX
XX
DT 15-JAN-2001 (first entry)
XX
DE A murine heparanase polypeptide.
XX
KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.
XX
XX Mus sp.
XX
XX W0200052178-A1.
XX
XX
PD 08-SEP-2000.
XX
PF 14-FEB-2000; 2000WO-US0003542.
XX
XX 01-MAR-1999; 99US-00258892.
XX
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX (FRIE/) FRIEDMAN M M.
XX
XX Pecker I, Vlodavsky I, Feinstein E;
XX
```

```
WPI: 2000-579289/54.
N-PSDB: AAY5081.
XX
PT New polynucleotides encoding a polypeptide having heparanase activity,
PT useful in wound healing and in gene therapy, particularly in treating
PT tumor, inflammation, autoimmunity, neurodegenerative diseases.
XX
PS Claim 22; Page 144-145; 152pp; English.
XX
XX The present sequence represents murine protein with heparanase catalytic
XX activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
XX particularly in treating tumour, inflammation or autoimmunity.
XX Particularly, the polynucleotide is useful in modulating the
XX bioavailability of heparin-binding growth factors, cellular responses to
XX heparin-binding growth factors (e.g. bFGF) and cytokines (e.g.
XX interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular
XX susceptibility to certain viral and some bacterial and protozoa
XX infections, or disintegration of neurodegenerative plaques. The
XX polynucleotide is also useful in wound healing (e.g. thermal, chemical or
XX radiation burns), and in the treatment of angiogenesis, restenosis,
XX atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
XX Strausler Syndrome or Creutzfeldt-Jakob disease), and some viral,
XX bacterial or protozoa infections
XX
SQ Sequence 535 AA;
Query Match      92.5%; Score 2590.5; DB 3; Length 535;
Best Local Similarity 92.5%; Pred. No. 1.2e-239;
Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;

QY 1 MLRPLLLWLRALRALTQGTAGTAPTDVVDLEFYTKEFLFQSVSPSFLSITIDASLAT 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MLR-LLLLLWLPGLAQAQAGTAPTDVVDLEFYTKEPLRSVSPSFLSITIDASLAT 59

QY 61 DRPFLFLGSPRLRALARGSPAYLRFGGTYKTDPLIPDPNKEPTSEERSYQSDNDIC 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 DRPFLFLGSPRLRALARGSPAYLRFGGTYKTDPLIPDPNKEPTSEERSYQSDNDIC 119

QY 121 GSRVSADVLRLQMEWPFQELLRLREYQREKFNSTYSRSSVDMLYSFACSKSLDLIFG 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 RSEPVSAVLRLQVEMPFQELLRLREYQREKFNSTYSRSSVDMLYSFACSKSLDLIFG 179

QY 181 LNALRTPLDRWNSSNAQLLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 LNALRTPLDRWNSSNAQLLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDFV 239

QY 241 ELHKLQSAFONAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 ELHKLQSAFONAKLYGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRIATK 299

QY 301 EDFLSSDVLDTFILSVQIKLVTKEMTPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 EDFLSSDVLDTFILSVQIKLVTKEMTPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 359

QY 361 KLGLSAQLGIEVVMRQVFFGAGNYHLVDENPEPLPDYWLSSLKPKLGVKPLMSRVKGP 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 KLGLSAQMGIEVVMRQVFFGAGNYHLVDENPEPLPDYWLSSLKPKLGVKPLMSRVKGP 419

QY 421 RSKLRVYLHCTNVYHPRYREGDLTYVLNLHNVTXHLKLPMPFSRPVDKYLKPFSGDG 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 RSKLRVYLHCTNVYHPRYQEGDLTYVLNLHNVTXHLKVPFPLFRKPDVYLLKPSGPG 479

QY 481 LLSKSVOLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPFASGYGFFVIRNAKIAACI 536
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 LLSKSVOLNGQILKMWDEQTLPALTEKPLPAGSALSVPFASGYGFFVIRNAKIAACI 535

RESULT 8
AAB07811
ID ABB07811 standard; protein; 535 AA.
XX
XX ABB07811;
XX
```

DT 03-JUL-2002 (first entry)  
XX Mouse heparanase sequence.  
DE Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;  
KW anti-protozoan; neuroprotective; heparin; mouse.  
XX OS Mus musculus.  
XX PH Location/Qualifiers  
FT Peptide  
FT 1..17  
FT /note= "putative signal peptide"  
FT Protein  
FT 18..535  
FT /note= "mature protein"  
XX US2002034810-A1.  
XX 21-MAR-2002.  
XX 16-AUG-2001; 2001US-00930218.  
XX 20-SEP-2000; 2000US-00666390.  
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;  
XX WPI; 2002-338926/37.  
XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful  
XX to treat various heparin-related disorders and the signal peptide is  
XX useful in production of membrane-targeted or secreted recombinant  
XX proteins.  
XX Disclosure; Fig 1a; 39pp; English.  
XX The invention relates to an isolated avian and reptile nucleic acid,  
XX encoding a polypeptide with heparanase catalytic activity. The signal  
XX peptide of the nucleic acid can be used to express membrane-associated or  
XX secreted proteins in heterologous expression systems. The encoded  
XX polypeptides can be used to prevent tumour angiogenesis, metastasis and  
XX invasion, and to intervene with pathologies associated with impaired  
XX heparin-binding growth factors, cellular responses to heparin-binding  
XX growth factors and cytokines, cell interaction with plasma lipoproteins,  
XX cellular susceptibility to viral, protozoa and bacterial infections or  
XX disintegration of neurodegenerative plaques. The present sequence  
XX represents a mouse heparanase protein sequence used in similarity studies  
XX Sequence 535 AA;  
Query Match 92.5%; Score 2590.5; DB 5; Length 535;  
Best Local Similarity 92.5%; Pred. No. 1.2e-239;  
Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;  
QY 1 MLRPLLLMLWGRALRATQGTACTPTKQVVDLEFYTQKLFQSVSPSFLSITIDASLAT 60  
DB 1 MLR-LLLLLWGLGALQAQAGACTPTDDVVDLEFYTQKPLRSVSPSFLSITIDASLAT 59  
QY 61 DPRFLTGLSPRLALARGISPAYLRGGTQTDLEIFDPNKEPTSEERSYQSQDNDIC 120  
DB 60 DPRFLTGLSPRLALARGISPAYLRGGTQTDLEIFDPNKEPTSEERSYQSQDNDIC 119  
QY 121 GSERVSADVLRLQKMEWPFQELLRLRQYQREFKNSTYSRSDMLYSFAKCSRLDIFG 180  
DB 120 RSEPVSAVLRKQGVVEFPQLLLRLRQYQKFNSTYSRSDMLYSFAKCSGLDIFG 179  
QY 181 LNALRTPDLRWNSNAQLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240  
DB 180 LNALRTPDLRWNSNAQLLNYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDFV 239  
QY 241 ELHKLLQKSAFQNAKLYGPDIGQPRGKTVKLLRSLFAGKAGGEVIDSLTWHHYLNGRVATK 300  
DB 240 ELHKLLQKSAFQNAKLYGPDIGQPRGKTVKLLRSLFAGKAGGEVIDSLTWHHYLNGRIATK 299

QY 301 EDFLSSDVLDTFILSVQKILKVTKEITPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 360  
DB 300 EDFLSSDVLDTFILSVQKILKVTKEITPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 359  
QY 361 KLGLSAQIGIEVVMRQVFFGAGNYHLVDENFEPLPDYMLSLFLKLVGPKVLMRSRVKGGPD 420  
DB 360 KLGLSAQMGIEVVMRQVFFGAGNYHLVDENFEPLPDYMLSLFLKLVGPKVLMRSRVKGGPD 419  
QY 421 RSKLRVYLHCTNVVHPRYREGDLTLYVLNLHNVTKHLKLPDMFSPDPVKYLLKPFSGDG 480  
DB 420 RSKLRVYLHCTNVVHPRYREGDLTLYVLNLHNVTKHLKLPDMFSPDPVKYLLKPFSGDG 479  
QY 481 LLSKSVQLNGOTLKMVDQOTLPALTEKPLPAGSSLSVPAPFSYGFVIRNAKIAACI 536  
DB 480 LLSKSVQLNGQILKMWDEQOTLPALTEKPLPAGSALSPLAFSYGFVIRNAKIAACI 535  
RESULT 9  
ADG88834  
ID ADG88834 standard; protein; 535 AA.  
XX AC ADG88834;  
XX DT 11-MAR-2004 (first entry)  
XX DE Mouse hpa protein.  
XX KW Wound healing; heparanase; ulcer; burn; laceration; surgical incision;  
XX KW necrosis; pressure wound; diabetic ulcer; angiogenesis; mouse; therapy.  
XX OS Mus musculus.  
XX PN US2003161823-A1.  
XX PD 28-AUG-2003.  
XX PF 14-JAN-2003; 2003US-00341582.  
XX PR 31-AUG-1998; 98WO-US017954.  
XX PR 01-MAR-1999; 99US-00258892.  
XX PR 06-FEB-2001; 2001US-00776874.  
XX PR 05-SEP-2001; 2001WO-IL000830.  
XX PR 19-NOV-2001; 2001US-00988113.  
XX PA (ILAN/) ILAN N.  
XX PA (VLOD/) VLODAVSKY I.  
XX PA (YACO/) YACOBY-ZEEVI O.  
XX PA (PECK/) PECKER I.  
XX PA (FEIN/) FEINSTEIN E.  
XX Ian N, Vlodavsky I, Yacoby-Zeevi O, Pecker I, Feinstein E;  
XX WPI; 2003-897910/82.  
XX N-FSDB; ADG88833, ADG88835.  
XX Composition for treating a wound comprising recombinant heparanase is  
XX useful to induce or accelerate wound healing and induce or accelerate  
XX angiogenesis.  
XX Claim 2; SEQ ID NO 44; 143pp; English.  
XX The present invention relates to methods and compositions for inducing  
XX and/or accelerating wound healing via the catalytic activity of  
XX heparanase. The invention is used to induce or accelerate a healing  
XX process, particularly of an ulcer, burn, laceration, surgical incision,  
XX necrosis, pressure wound, diabetic ulcer and to induce or accelerate  
XX angiogenesis. The present sequence is mouse hpa protein.  
XX Sequence 535 AA;  
Query Match 92.5%; Score 2590.5; DB 7; Length 535;  
Best Local Similarity 92.5%; Pred. No. 1.2e-239;



Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;	
QY	1 MLRPLLLMLWGRRLALTOGTAGTAPTKOVVDLEFYTKRLFQSVSPSFLSITIDASLAT 60
Db	1 MLR-LLLLMLWGLGALAQAGAPAGTAPTDVVDLEFYTKRPLRSVSPSFLSITIDASLAT 59
QY	61 DRPFLTLGSPRLRALARGLSPAYLRFGGTTDFLIIDPNKPTSEERSYQSQDNNDIC 120
Db	60 DRPFLTLGSPRLRALARGLSPAYLRFGGTTDFLIIDPNKPTSEERSYQSQVNHIDC 119
QY	121 GSRVSAVLRKLQMEWPFQELLRLREYQREFKNSTYSRSSVDMLYSFACSKRLDLIFG 180
Db	120 RSEPVSAVLRKLQVWPFQELLRLREYQREFKNSTYSRSSVDMLYSFACSKGLDLIFG 179
QY	181 LNALRTDPLRWNSNAQLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
Db	180 LNALRTDPLRWNSNAQLLNYCSSKGYNISWELGNEPNSFWKKAHLIDGLQGEDFV 239
QY	241 ELHKLQSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300
Db	240 ELHKLQSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRIATK 299
QY	301 EDFLSSDALDTFILSVQKILKVTKETPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 360
Db	300 EDFLSSDALDTFILSVQKILKVTKETPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 359
QY	361 KLGLSAQLGIEVVMRQVFFGAGNYHLVDENFELPDYWLKLLPKLVGPKVMSRVKGP 420
Db	360 KLGLSAQMGIIEVVMRQVFFGAGNYHLVDENFELPDYWLKLLPKLVGPRVLSRVKGP 419
QY	421 RSKLRVYLHCTNVYHPRYREGDLTLVYLNHNVTYKHLKLPMPFSRPVDKYLLKPFSGDG 480
Db	420 RSKLRVYLHCTNVYHPRYREGDLTLVYLNHNVTYKHLKVPDPFRKPDVYLLKPSGPDG 479
QY	481 LLSKSVOLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPAFSYGFVIRNAKIAACI 536
Db	480 LLSKSVOLNGQILKMWDEQTLPALTEKPLPAGSALSIPAFSYGFVIRNAKIAACI 535
RESULT 10	
ID	ADLI6413
XX	ADLI6413 standard; protein; 535 AA.
AC	ADLI6413;
XX	
DT	06-MAY-2004 (first entry)
XX	
DE	Mouse heparanase protein.
XX	
KW	Mouse; heparanase; enzyme; heparanase-dependent cancer; cancer;
KW	autoimmune reaction; inflammation.
XX	
OS	Mus musculus.
XX	
PN	US2003236215-A1.
XX	
PD	25-DEC-2003.
XX	
PF	09-JUN-2003; 2003US-00456573.
XX	
PR	31-AUG-1998; 98WO-US017954.
PR	01-MAR-1999; 99US-00258892.
PR	08-NOV-1999; 99US-00435739.
XX	
PA	(INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA	(HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX	
PI	Pecker I, Vlodavsky I, Feinstein E;
XX	
XX	WPI; 2004-070610/07.
XX	
PT	New antisense oligonucleotide hybridizable with a polynucleotide encoding
PT	a polypeptide with heparanase activity, useful for treating diseases such

PT	as cancer and autoimmune disorders.
XX	
PS	Claim 3; SEQ ID NO 44; 108pp; English.
XX	
CC	The invention relates to an antisense oligonucleotide (ASO) comprising a
CC	polynucleotide or a polynucleotide analogue of at least 10 bases being
CC	hybridisable in vivo , under physiological conditions, with a portion of
CC	a polynucleotide strand encoding a polypeptide having heparanase
CC	catalytic activity. Also included are a method of in vivo downregulating
CC	heparanase activity (comprising administering the ASO in vivo), a method
CC	of treating a subject suffering from a pathological condition
CC	(characterised by heparanase activity, comprising administering ASO to
CC	the subject), a pharmaceutical composition comprising the ASO and a
CC	carrier, an antisense nucleic acid construct (comprising a promoter
CC	sequence and a polynucleotide sequence directing the synthesis of an
CC	antisense RNA sequence of at least 10 bases being hybridisable in vivo ,
CC	under physiological conditions, with a polynucleotide strand encoding a
CC	polypeptide having heparanase catalytic activity), a method of in vivo
CC	downregulating heparanase activity (comprising administering in vivo the
CC	antisense nucleic acid construct), a pharmaceutical composition
CC	comprising the antisense nucleic acid construct and a carrier, and an
CC	antisense oligonucleotide comprising a polynucleotide or a polynucleotide
CC	analogue of at least 10 bases being hybridisable in vivo , under
CC	physiological conditions, with a portion of a polynucleotide strand being
CC	characterised by forming at least a portion of an untranslated region
CC	(UTR) for a polynucleotide strand encoding a polypeptide having
CC	heparanase catalytic activity. The methods and compositions of the
CC	present invention are useful for the prevention and/or treatment of
CC	diseases or conditions associated with aberrant heparanase activity, such
CC	as heparanase-dependent cancer, cancer, autoimmune reaction and
CC	inflammation. The gene for human heparanase is located on chromosome 4.
CC	The present sequence is the mouse heparanase protein.
XX	
QQ	Sequence 535 AA;

Query Match 92.5%; Score 2590.5; DB 8; Length 535;  
Best Local Similarity 92.5%; Pred. No. 1.2e-239;  
Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;

QY	1 MLRPLLLMLWGRRLALTOGTAGTAPTKOVVDLEFYTKRLFQSVSPSFLSITIDASLAT 60
Db	1 MLR-LLLLMLWGLGALAQAGAPAGTAPTDVVDLEFYTKRPLRSVSPSFLSITIDASLAT 59
QY	61 DRPFLTLGSPRLRALARGLSPAYLRFGGTTDFLIIDPNKPTSEERSYQSQDNNDIC 120
Db	60 DRPFLTLGSPRLRALARGLSPAYLRFGGTTDFLIIDPNKPTSEERSYQSQVNHIDC 119
QY	121 GSRVSAVLRKLQMEWPFQELLRLREYQREFKNSTYSRSSVDMLYSFACSKRLDLIFG 180
Db	120 RSEPVSAVLRKLQVWPFQELLRLREYQREFKNSTYSRSSVDMLYSFACSKGLDLIFG 179
QY	181 LNALRTDPLRWNSNAQLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
Db	180 LNALRTDPLRWNSNAQLLNYCSSKGYNISWELGNEPNSFWKKAHLIDGLQGEDFV 239
QY	241 ELHKLQSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK 300
Db	240 ELHKLQSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRIATK 299
QY	301 EDFLSSDALDTFILSVQKILKVTKETPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 360
Db	300 EDFLSSDALDTFILSVQKILKVTKETPGKKVWLGETSSAYGGAPLLSNTFAAGFMWLD 359
QY	361 KLGLSAQLGIEVVMRQVFFGAGNYHLVDENFELPDYWLKLLPKLVGPKVMSRVKGP 420
Db	360 KLGLSAQMGIIEVVMRQVFFGAGNYHLVDENFELPDYWLKLLPKLVGPRVLSRVKGP 419
QY	421 RSKLRVYLHCTNVYHPRYREGDLTLVYLNHNVTYKHLKLPMPFSRPVDKYLLKPFSGDG 480
Db	420 RSKLRVYLHCTNVYHPRYREGDLTLVYLNHNVTYKHLKVPDPFRKPDVYLLKPSGPDG 479
QY	481 LLSKSVOLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPAFSYGFVIRNAKIAACI 536



PA (SHLO/) SHLOMI Y.  
PA (PECK/) PECKER I.  
PA (AYAL/) AYAL-HERSHKOVITZ M.  
PA (FEIN/) FEINSTEIN E.  
PA (VGEL/) VAN GELDER J M.  
PA (VLOD/) VLODAVSKY I.  
PA (FRIE/) FRIEDMANN Y.  
XX  
PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;  
PI Friedmann Y;  
XX  
DR WPI; 2004-625084/60.  
XX  
PT Targeted drug delivery to a heparanase-expressing tissue of a patient,  
PT useful for treating heparanase-associated conditions such as inflammation  
PT or cancer, comprises administering a drug and an anti-heparanase antibody  
PT complex.  
XX  
PS Claim 2; SEQ ID NO 2; 58pp; English.  
XX  
CC The invention relates to a method of targeted drug delivery to a tissue  
CC of a patient, the tissue expressing heparanase. The method comprises  
CC providing a complex of a drug directly or indirectly linked to an anti-  
CC heparanase antibody, and administering the complex to the patient. In the  
CC targeted drug delivery, the antibody comprises an epitope or its portion  
CC capable of specifically binding to at least one epitope of a heparanase  
CC protein. The composition and methods of the invention are useful for  
CC diagnosing, preventing or treating conditions associated with heparanase  
CC catalytic activity (e.g. an inflammatory disorder, wound, scar,  
CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell  
CC proliferation, invasion of circulating tumour cells and metastatic  
CC disease), for purifying heparanase, or for developing drugs for those  
CC heparanase-associated conditions. The vasculopathy is atherosclerosis,  
CC restenosis or aneurysm. The cancerous condition is a solid cancer or a  
CC non-solid cancer. The solid cancer is a haematopoietic malignancy  
CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous  
CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous  
CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,  
CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and  
CC multiple myeloma. The solid cancer is selected from tumours in lip and  
CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,  
CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,  
CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of  
CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue  
CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,  
CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic  
CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary  
CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,  
CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,  
CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,  
CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's  
CC sarcoma. The present sequence is mouse heparanase.  
XX  
SQ Sequence 535 AA;

Query Match 92.5%; Score 2590.5; DB 8; Length 535;  
Best Local Similarity 92.5%; Pred. No. 1.2e-239;  
Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;

Db 180 LNALRTPDLRWNSSNAQLLLDYCSSKGYNISWELGNEPNSFWKKAHLIDGLQGEDFV 239  
Qy 241 ELHKLLQKSAFQNAKLYGPDIGQPRGKTIVKLLRSFLKAGGEVIDSLTWHHYLLNGRVATK 300  
Db 240 ELHKLLQKSAFQNAKLYGPDIGQPRGKTIVKLLRSFLKAGGEVIDSLTWHHYLLNGRIATK 299  
Qy 301 EDFLSSDVLDTFFILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLD 360  
Db 300 EDFLSSDALDTFFILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLD 359  
Qy 361 KLGLSAQLGIEVVMQVFFGAGNVLVDENPELPDYWLSLLFKLVGPKVLMRSVRKGD 420  
Db 360 KLGLSAQMGIEVVMQVFFGAGNVLVDENPELPDYWLSLLFKLVGPKVLMRSVRKGD 419  
Qy 421 RSKLRVYLHCTNVYHPRYREGDLTLVYVNLHNVTKHLKLPMPFSRPVDKYLKPFPGSDG 480  
Db 420 RSKLRVYLHCTNVYHPRYQEGDLTLVYVNLHNVTKHLKVPPLPRKPDVTYLLKPSGPDG 479  
Qy 481 LLSKSVOLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPAFSYGFVIRNAKIAACI 536  
Db 480 LLSKSVOLNGQILKMWDEQTLPALTEKPLPAGSALSPLAFSYGFVIRNAKIAACI 535  
RESULT 13  
ADT78175  
ID ADT78175 standard; protein; 535 AA.  
XX  
AC ADT78175;  
XX  
DT 13-JAN-2005 (first entry)  
XX  
DE Mouse heparanase protein.  
XX  
KW Antibody; epitope; heparanase; pathological condition; angiogenesis;  
KW cell proliferation; cancerous condition; tumour cell invasion;  
KW metastatic disease; heparanase-related disorder; inflammatory disorder;  
KW wound; scar; vasculopathy; autoimmune condition; renal disease;  
KW cytostatic; antiinflammatory; vulneryary; antiarteriosclerotic;  
KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; mouse.  
XX  
OS Mus musculus.  
XX  
FH Key Location/Qualifiers  
FT Binding-site 149..154 /note= "Putative heparin binding site"  
FT Binding-site 263..269 /note= "Putative heparin binding site"  
FT Binding-site 419..425 /note= "Putative heparin binding site"  
XX  
PN US2004213789-A1.  
XX  
PD 28-OCT-2004.  
XX  
XX 22-AUG-2003; 2003US-00645659.  
XX  
XX 02-SEP-1997; 97US-00922170.  
PR 01-MAY-1998; 98US-00071739.  
PR 04-NOV-1998; 98US-00186200.  
PR 19-FEB-2003; 2003US-00368044.  
XX  
PA (YACO/) YACOBY-ZEEVI O.  
PA (PERE/) PERETZ T.  
PA (MIRO/) MIRON D.  
PA (SHLO/) SHLOMI Y.  
PA (PECK/) PECKER I.  
PA (AYAL/) AYAL-HERSHKOVITZ M.  
PA (FEIN/) FEINSTEIN E.  
PA (VGEL/) VAN GELDER J M V.  
PA (VLOD/) VLODAVSKY I.  
PA (FRIE/) FRIEDMANN Y.  
XX

PI Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;  
PI Friedmann Y;  
XX  
XX  
DR WPI; 2004-774790/76.

XX New neutralizing monoclonal anti-heparanase antibodies, useful for  
PT detecting, treating or preventing cancer, inflammatory or autoimmune  
PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.  
XX  
XX  
PS Claim 5; SEQ ID NO 2; 68pp; English.

XX The invention relates to an isolated antibody or antibody portion capable  
XX of specifically binding to or elicited by at least one epitope of a  
XX heparanase protein, where the heparanase protein is at least 60%  
XX homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and  
XX where at least one epitope comprises a sequence at least 70% homologous  
XX to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)  
XX a hybridoma cell line comprising a cell line for producing the monoclonal  
XX antibody, (b) a method for detecting, treating or preventing a  
XX pathological condition or a heparanase-related disorder or condition in a  
XX subject, (c) a method for monitoring the state of a heparanase-related  
XX disorder or condition in a subject, and (d) a pharmaceutical composition  
XX comprising the isolated anti-heparanase antibody or antibody portion and  
XX a pharmaceutical carrier. The antibody, methods, and composition are  
XX useful for detecting, treating, preventing or monitoring a pathological  
XX condition, e.g. angiogenesis, cell proliferation, a cancerous condition  
XX (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,  
XX or prostate cancer), minor cell proliferation, invasion of circulating  
XX tumour cells, or a metastatic disease, or a heparanase-related disorder  
XX or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy  
XX (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or  
XX renal disease or disorder (diabetic nephropathy, glomerulosclerosis,  
XX nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell  
XX carcinoma) in a mammal. This sequence represents mouse heparanase.

XX Sequence 535 AA;

Query Match 92.5%; Score 2590.5; DB 8; Length 535;  
Best Local Similarity 92.5%; Pred. No. 1.2e-239;  
Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;

QY 1 MLRPLLLWLGRLALTOGTPAGTAPKDVLDLEFYTKRLFQSVSPSFLITIDASLAT 60  
DB 1 MLR-LLLLWLGPLGALQAQAPAGTAPTDVLDLEFYTKRPLRSVPSFLITIDASLAT 59  
QY 61 DPLFLPLGSPRLALARGLSPAYLRGGTKTDFLIIDPNKEPTSEERSYQSQDNDIC 120  
DB 60 DPLFLPLGSPRLALARGLSPAYLRGGTKTDFLIIDPNKEPTSEERSYQSQVNHIDIC 119  
QY 121 GSERVSADVLRLQWMPFOELLILLRQYQREFKNSTYSSVDMLYSFAKCSRLDLIFG 180  
DB 120 RSEPVSAVLRKLQWMPFOELLILLRQYQREFKNSTYSSVDMLYSFAKCSGLDLIFG 179  
QY 181 LNALLRTPDLRWSSNNAQLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240  
DB 180 LNALLRTPDLRWSSNNAQLLNYCSSKGYNISWELGNEPNSFWKKAHLIDGLQGEDFV 239  
QY 241 ELHKLQKSAFONAKLYGPDIGQPRGKTVKLLRSFLKAGGEVDLSLTWHYHLYNGRVATK 300  
DB 240 ELHKLQKSAFONAKLYGPDIGQPRGKTVKLLRSFLKAGGEVDLSLTWHYHLYNGRIATK 299  
QY 301 EDFLSSDVLDTFLSVQKILKVTKEWTPGKVMWLGETSAYGGCAPLLSNTFAAGFWMLD 360  
DB 300 EDFLSSDVLDTFLSVQKILKVTKEWTPGKVMWLGETSAYGGCAPLLSNTFAAGFWMLD 359  
QY 361 KGLGSAQLGIEVWNRQVFFGAGNHLVDENFEPPLDYWLKLLFKLVGPKVLMRSVRKQPD 420  
DB 360 KGLGSAQLGIEVWNRQVFFGAGNHLVDENFEPPLDYWLKLLFKLVGPKVLMRSVRKQPD 419  
QY 421 RSKLRVYLHCTNVYHPYRREGDLYVNLNHNVTKHLKLPMPFMRPVDKYLLKPFQSDG 480  
DB 420 RSKLRVYLHCTNVYHPYRREGDLYVNLNHNVTKHLKLPMPFMRPVDKYLLKPFQSDG 479

QY 481 LLSKSVQLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPAFSYGFFVIRNAKTAACI 536  
DB 480 LLSKSVQLNGQTLKMWDEQTLPALTEKPLPAGSALSPLAFSYGFFVIRNAKTAACI 535

RESULT 14  
AEA42424

ID AEA42424 standard; protein; 535 AA.

XX AC AEA42424;

XX DT 28-JUL-2005 (first entry)

XX DE Mouse heparanase epitope peptide SEQ ID NO:2.

XX KW antibody; heparanase; antiinflammatory; vulnary; immunosuppressive;  
XX KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;

XX KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;  
XX KW angiogenesis disorder; cancer; tumor; metastasis.

XX OS Mus musculus.

XX PN AU2004201462-A1.

XX PD 06-MAY-2004.

XX PF 08-APR-2004; 2004AU-00201462.

XX PR 08-APR-2004; 2004AU-00201462.

XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

PI Vlodavsky I, Pecker I, Miron M, Gilboa A, Miron D, Moskowitz H;  
PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;  
PI Feinstein E;

XX WPI; 2005-173343/19.

PT Novel isolated antibody capable of specifically binding to epitope of  
PT heparanase protein, useful for preventing and treating heparanase-related  
PT disorder such as inflammatory disorder, scars, autoimmune conditions or  
XX angiogenesis.

PS Claim 2; SEQ ID NO 2; 260pp; English.

XX The invention relates to an isolated antibody or its portion (I) capable  
XX of specifically binding to an epitope of a heparanase protein. Also  
XX described: (1) a cell line (II) for producing a monoclonal antibody or  
XX its portion, comprising a cell line for producing (I); (2) a  
XX pharmaceutical composition comprising (I) and a carrier; and (3) an  
XX affinity medium (III) for binding human heparanase polypeptides,  
XX comprising (I) immobilized to a chemically inert, insoluble carrier. (1)  
XX useful for treating a subject suffering from a pathological condition,  
XX which involves administering (I) to the subject. (I) is useful for  
XX preventing and treating heparanase-related disorder or condition chosen  
XX from inflammatory disorder, wound, scar, vasculopathy, autoimmune  
XX condition, angiogenesis, cell proliferation, cancerous condition, tumor  
XX cell proliferation, invasion of circulating tumor cells and metastatic  
XX disease. (I) is useful for detecting the presence of heparanase  
XX polypeptide in a sample. (I) is useful for detecting heparanase-related  
XX disease or condition in a subject such as vertebrate, preferably mammal  
XX e.g., human. The heparanase-related disorder or condition further  
XX includes renal disease or disorder chosen from diabetic nephropathy,  
XX glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome  
XX and renal cell carcinoma. The present sequence represents mouse  
XX heparanase, which is used in the exemplification of the present  
XX invention.

XX Sequence 535 AA;

Query Match 92.5%; Score 2590.5; DB 9; Length 535;

Best Local Similarity 92.5%; Pred. No. 1.2e-239;		Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;	
QY	1	MLRPLLLMLWGRRLALTOCTPAGTAPTKDQVLDLEFFYTKELFQSPSPFLSITIDASLAT	60
DB	1	MLR-LLLLMLWGLGALAQAGPAGTAPDDVDLEFFYTKPLRSVSPFLSITIDASLAT	59
QY	61	DRPFLFLGSPRLRALARGLSPAYLRFGTKTDFLIPDPNKEPTSEERSYQSQDNDIC	120
DB	60	DRPFLFLGSPRLRALARGLSPAYLRFGTKTDFLIPDPNKEPTSEERSYQSQVNHIC	119
QY	121	GSESVADVLRKLQMEWPFQELLRLLEQYQREPKNSYSSSDVMLYSPAKCSRLDLIFG	180
DB	120	RSEPVSAVLRKLQMEWPFQELLRLLEQYQREPKNSYSSSDVMLYSPAKCSGLDLIFG	179
QY	181	LNALLRTPDLRWNSSNAQLLNTYSSKGYNISWELGNEPNSFWKKAQISIDGLQLGDFV	240
DB	180	LNALLRTPDLRWNSSNAQLLNTYSSKGYNISWELGNEPNSFWKKAHILIDGLQLGDFV	239
QY	241	ELHKLQSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVATK	300
DB	240	ELHKLQSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRIATK	299
QY	301	EDPLSSDVLDTFILSVQKILKVTKEMTPGKQWLGWGETSSAYGGAPLLSNTFAAGFWMLD	360
DB	300	EDPLSSDALDTFILSVQKILKVTKETTPGKQWLGWGETSSAYGGAPLLSNTFAAGFWMLD	359
QY	361	KLGLSAQLGIEVVMROVFFGAGNYHLVDENFELPDYWLSSLFKKLGPVKLMSRVKGP	420
DB	360	KLGLSAQMGIEVVMROVFFGAGNYHLVDENFELPDYWLSSLFKKLGPVKLMSRVKGP	419
QY	421	RSKLRVYLHCTNHYHPRYREGDLTYVNLNHNVTKHLKLPMPMFSRPPVDKYLKLPFGSDG	480
DB	420	RSKLRVYLHCTNHYHPRYQEGDLTYVNLNHNVTKHLKVPPLFRKPDVITYLLKPSGPDG	479
QY	481	LLSKSVOLNGQTLKMVDEQTLPALTEKPLPAGSSLSVPAFSYGFFVIRNAKIAACI	536
DB	480	LLSKSVOLNGQILKMVDEQTLPALTEKPLPAGSALSVPAFSYGFFVIRNAKIAACI	535
RESULT 15			
ID	ADY27034	standard; protein; 545 AA.	
AC	ADY27034;		
XX	05-MAY-2005 (first entry)		
DE	Bovine heparanase protein.		
KW	Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;		
KW	neurological disease; viral infection; infection; cystostatic;		
KW	antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;		
XX	protease; enzyme; enzyme purification.		
OS	Bos taurus.		
XX	WO2005016227-A2.		
XX	24-FEB-2005.		
XX	12-AUG-2004; 2004WO-IL000744.		
XX	14-AUG-2003; 2003US-0494800P.		
PR	12-JAN-2004; 2004US-0535492P.		
XX	(INSI-) INSIGHT BIOPHARMACEUTICALS LTD.		
XX	Van-Gelder JM, Miron D;		
XX	WPI; 2005-182203/19.		
PT	Regulating heparanase activity, useful for treating heparanase-associated		

PT	diseases (e.g. cancer, inflammation, cardiovascular diseases,		
PT	neurological diseases or viral diseases) comprises modulating heparanase		
PS	activation.		
XX	Disclosure; SEQ ID NO 6; 211pp; English.		
XX	The invention relates to a method of regulating heparanase activity in a		
CC	tissue or regulating a biological process depending at least in part on		
CC	heparanase activity comprising modulating heparanase activation. The		
CC	invention also relates to methods of treating a heparanase- or heparin		
CC	binding protein-associated disease or disorder in a subject, a		
CC	pharmaceutical composition for use in the treatment of a heparanase-		
CC	associated disease or disorder comprising a therapeutic amount of an		
CC	agent capable of modulating heparanase activation and a pharmaceutical		
CC	carrier or diluent, a method of identifying a protease activator of		
CC	heparanase, a protease substrate mimetic comprising a peptide		
CC	representing a subset or all substrate residues or cleavage sites of		
CC	human heparanase or an equivalent non-human heparanase, a method of		
CC	producing active heparanase and a method of modulating an adhesion		
CC	activity of heparanase. The composition and methods are useful for		
CC	modulating heparanase activation and for treating heparanase-associated		
CC	diseases or disorders such as cancer, inflammation, cardiovascular		
CC	diseases, neurological diseases or viral infections. This sequence		
CC	represents a bovine heparanase protein used in the scope of the		
CC	invention.		
XX	Sequence 545 AA;		
QY	Query Match 76.8%; Score 2151; DB 9; Length 545;		
DB	Best Local Similarity 76.4%; Pred. No. 2.5e-197;		
DB	Matches 410; Conservative 50; Mismatches 75; Indels 2; Gaps 1;		
QY	2	LRP--LLLLMLWGRRLALTOCTPAGTAPTKDQVLDLEFFYTKELFQSPSPFLSITIDASIA	59
DB	9	LRPLLLLLLPLGLPGCSGTPAAAAPADDAELEFFTERPLHLVSPAFSLFTIDANLA	68
QY	60	TDRPFLFLGSPRLRALARGLSPAYLRFGTKTDFLIPDNKEPTSEERSYQSQDNDI	119
DB	69	TDRPFLFLGSSKLRTLARGLAPAYLRFGNGKGDFLIFDPKKEPAFEERSYWLQSNDI	128
QY	120	CGSERVADVLRKLQMEWPFQELLRLLEQYQREPKNSYSSSDVMLYSPAKCSRLDLIF	179
DB	129	CKSGSIPSDVEEKLRLQWPFQELLRLLEQYQREPKNSYSSSDVMLYSPAKCSRLDLIF	188
QY	180	GLNALLRTPDLRWNSSNAQLLNTYSSKGYNISWELGNEPNSFWKKAQISIDGLQLGDF	239
DB	189	GVNALLRTPDLRWNSSNAQLLNTYSSKGYNISWELGNEPNSFWKKAQIFINGQLGDF	248
QY	240	VELHKLQSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHHYLNGRVAT	299
DB	249	IEPRKLQSAFQNAKLYGPDIGQPRENTVMKLSFLKAGGEVIDSVTWHHYVNGRIAT	308
QY	300	KEPFLSSDVLDTFILSVQKILKVTKEMTPGKQWLGWGETSSAYGGAPLLSNTFAAGFWML	359
DB	309	KEDFLNPDIIDTFTISSVQKTLRIVEKIRPLKQWLGWGETSSAFGGAPPLSNTFAAGFWML	368
QY	360	DKLGLSAQLGIEVVMROVFFGAGNYHLVDENFELPDYWLSSLFKKLGPVKLMSRVKGP	419
DB	369	DKLGLSARMGIEVVMROVFFGAGNYHLVDGNFELPDYWLSSLFKKLGPVKLMSRVKGP	428
QY	420	DRSKLRVYLHCTNHYHPRYREGDLTYVNLNHNVTKHLKLPMPMFSRPPVDKYLKLPFGSD	479
DB	429	DRSKFRVYLHCTNHYHPRYKEGDLTYVNLNHNVTKHLKLPMPMFSRPPVDKYLKLPFGSD	488
QY	480	GLLSKSVOLNGQTLKMVDEQTLPALTEKPLPAGSSLSVPAFSYGFFVIRNAKIAACI	536
DB	489	GLLSKSVOLNGQILKMVDEQTLPALTEKPLPAGSSLSVPAFSYGFFVIRNAKIAACI	545
Search completed: June 5, 2006, 12:09:42			
Job time : 107.86 secs			

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:10:07 ; Search time 17.8454 Seconds  
(without alignments)  
2889.939 Million cell updates/sec

Title: US-10-645-659A-3  
Perfect score: 2800  
Sequence: 1 MLRPLLLWLWGLRALTQG.....VPAFSYGFVIRNAKIAACI 536

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: Piri:\*  
2: Piri2:\*  
3: Piri3:\*  
4: Piri4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	904	32.3	480	2 JC7506	heparanase protein
2	410	14.6	521	2 T45608	hypothetical prote
3	162.5	5.8	190	2 T01953	hypothetical prote
4	109	3.9	575	2 T45668	hypothetical prote
5	109	3.9	1331	2 A48954	mannan endo-1,4-be
6	108.5	3.9	4436	2 E71086	hypothetical prote
7	108	3.9	725	2 S01042	aerobactin recepto
8	107	3.8	837	1 A31842	endo-1,4-beta-xyla
9	106.5	3.8	1573	2 T50113	3-dehydroquinatase
10	106	3.8	599	2 S75363	hypothetical prote
11	106	3.8	2165	1 RN2A2	genome polyprotein
12	105.5	3.8	575	2 T12094	beta-fructofuranos
13	104.5	3.7	1482	2 S13495	pregnancy zone pro
14	104	3.7	511	2 S61166	probable membrane
15	104	3.7	732	2 T44483	receptor-like prot
16	104	3.7	796	2 D97065	transketolase [imp
17	103.5	3.7	844	2 T52396	formin-binding pro
18	103.5	3.7	914	2 B96592	hypothetical prote
19	103.5	3.7	1462	2 T06819	DNA topoisomerase
20	103	3.7	361	2 A43510	basic membrane pro
21	102.5	3.7	587	2 S6231	beta-fructofuranos
22	102	3.6	879	2 F86875	probable fibrial
23	102	3.6	879	2 YG1031	probable outer mem
24	102	3.6	1392	1 EGBYAD	L-aminoadipate-sem
25	101.5	3.6	412	2 G81581	tyrosyl-tRNA synth
26	101	3.6	581	2 E90449	conserved hypotet
27	101	3.6	596	2 T04506	hypothetical prote
28	101	3.6	709	2 A35364	carcinoembryonic a
29	100.5	3.6	495	2 C89778	hypothetical prote

RESULT 1

JC7506

heparanase protein 2a - human

C:Species: Homo sapiens (man)

C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 09-Jul-2004

C:Accession: JC7506

R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hircock, M.; Pat

Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000

A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family me

A:Reference number: JC7506

A:Accession: JC7506

A:Molecule type: mRNA

A:Residues: 1-480 <MCK>

A:Cross-references: UNIPROT:Q9HB39; UNIPARC:UPI000003E88A; GB:AF282885

C:Comment: This protein, an intracellular membrane-bound enzyme, has biological and therape

C:Genetics:

A:Gene: hpa2a

A:Map position: 10q23-10q24

C:Keywords: heparin binding; membrane bound

Query Match 32.3%; Score 904; DB 2; Length 480;

Best Local Similarity 37.8%; Pred. No. 6.1e-62;

Matches 202; Conservative 74; Mismatches 150; Indels 108; Gaps 8;

Qy 12 GRLRALTQGTAGTAPTADVVDLEFYTKRLFSQSPSFLSITIDASLATDPRFLTLGSP 71

Db 42 GDRPLPVDRAAG-LKEKTLILLDVSTKNPVTNENFLSLQLDPSIIHD-GWLDLSSK 99

Qy 72 RLRLARGLSPAYLRFQGTKTDFLIDFNKPTSEERSYWSQDNNDICGSERVSADVL 131

Db 100 RLVTLARGLSPAFLRFGGKRTDFLQF----- 125

Qy 132 KLQWENPQELLLEBEQYQREFKNSTYSRSSVDMLYSFAKCSRLDLIFGLNALLRTPDLR 191

Db 126 -----QNLNPAKRGSGG-----PD-- 141

Qy 192 WNSSNAQLLLNYCSSKGYNISWELGNENPSFWKKAQISIDGLQLGEDFVELHKLQK-SA 250

Db 142 -----YLLKYY-----EDENNYVTMHGRAVNSQLGKDYIQLKSLQPIRI 183

Qy 251 FQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRVATKPEFLSSDVL 310

Db 184 YSRASLYGPNIGRPRKNVIALLDGPMKVGASTDAVTWQHCVIDGRVVKVMDFLKRLLD 243

Qy 311 TFIILSVQKILVKTEMTPGKVKWLGETSSAVGGGAPLISNTFAAGFMWLDKGLSLQGL 370

Db 244 TLDQIRKIQKVVNTYTPGKKIWLGEVVTTSAGGTNNLSDSYAAGFLNLTIGMLANO 303

Qy 371 EVVMRQVFFGAGNVLVDENFEPLPDYWLISLLFKLVGPKVLMRSVKGPD-----R 421

Db 304 DVIRHSFFDHGYNHLVDQNFENPLPDYWLISLLFKLVGPKVLMRSVKGPD-----R 363

ALIGNMENTS



```

Qy 422 SKLRVYLHCTNVYHPYREGDITLVVLNLHNVYTKHLKLPMPNFSRPVDKYLLKPFQSDGL 481
Db 364 DKLRIYAHCTNHHNHNHYRGSITFLIINLHRGRKKIKLAGTLDKLVHQYLLQPYQEGSL 423
Qy 482 LSKSVQLNGOTLKMVDEQTLPALTEKPLPAGSSLSVPAFSYGFVFI RNAKIAAC 535
Db 424 KSKSVQLNGQPLVMVDDGTLPKLPKPRFRAGRTLVIPPTMGVFFVYKVNALAC 477

RESULT 2
T45608
hypothetical protein F13G24.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C:Accession: T45608
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckae
submitted to the Protein Sequence database, December 1999
A:Reference number: Z23009
A:Accession: T45608
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-521 <BEV>
A:Cross-references: UNIPROT:O9SDAL; UNIPARC:UPI00000AA497C; EMBL:AL133421
A:Experimental source: cultivar Columbia; BAC clone F13G24
C:Genetics:
A:Map position: 5
A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3
A:Note: F13G24.30

```

Query Match	14.6%	Score 410;	DB 2;	Length 521;
Best Local Similarity	29.4%	Pred. No. 1.1e-23;		
Matches	155;	Conservative	73;	Mismatches 179;
			Indels 120;	Gaps 25;

  

Qy	68	LGSPRLRALARGLSPAYLRFGGTXKTDFLIPDPNKEPTSEERSYWQSQDNNDICGSESVSA	127
Db	55	LTRPELLTKAKAFPLRIRIGGSLODQYVDVGNLKT	94
Qy	128	DVLRFKLQMWPPQBELLLREYQREFKNS---TYSRSV-----DMLYSFAKCSRDLILF	179
Db	95	-----PFOKM-----NSGLGFSGKCLHMKRWDELNSPLTATGAVVTF	132
Qy	180	GLNALRLTPDLR-----WNSSNAQLLLNYCSSKGYNI-SWELGNPNPNSFWKKAQISID	231
Db	133	GLNALRGHKLGRKAWGANDHINTQDFLNTVTSKGVIDSWEFGNELSGSGVGASVSAE	192
Qy	232	GLQGEDEFVELHKLLOKSAFONAKLPGDIDGPRG-----KTVLLRSFLKAGEVIDSL	286
Db	193	--LYGKOLIVLKDINK-VYKNSWLHKPILVAPGFGVEQOQWYTKLLET---SGSVVDVV	246
Qy	287	TWHYYLNGRVAEKEDFLSSDVLD--TFILSVQIKL-----VTKEMTPEKKVWLGETSSAY	341
Db	247	T-HHIYNLG--SGNDPALVKKIMPSVLSQVSKTFKDVNQTQIEHGWPASWPWGSEGGAY	303
Qy	342	GGGAPLLSNTFAAGFMWLDKLGLSAQLGIEVVMQVFFGAGNYHLVDE-NFEPLPDYWL	400
Db	304	NSGGRHVSDFIDSFYWLQDQIGMSARHNTKVYCRQTLVG-GFYGLLEKGTVPNPDYSA	362
Qy	401	LLFKKLGPVKVLMRSKVRKPDRLVYLHCTNVYHPRYREGDGLTYLVNLHN-----V	453
Db	363	LLWHRLMGKVLAVQTDGP--POLRVYAHCS-----KBRAG-VTLLLLINLSNQSDFTVSV	414
Qy	454	TKHLKLPMPMPSRPVDKYL-----KPF-----GSDGLL-----SKSVQL	488
Db	415	SGINNVVLNAESRK-KKSLDLTKRPFSWIGSKASDGYLAREEYHLTPENGVLRSKTMVL	473
Qy	489	NGOTLKMWDEQTLPALTEKPLPAGSSLSVPAFSYGFFVIRNAKAAAC	535
Db	474	NGKSLKATATGDIPLSEPLVRSVNSPLNVLPLMSFVLNPFNDASAC	520

  

RESULT 3	
T01953	

RESULT 3  
T01953

hypothetical protein T2L5.6 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 09-Jul-2004  
R:Geisel, C.; Smith, A.; Le, T.  
submitted to the ENBL Data Library, October 1998  
A:Description: The sequence of A. thaliana T2L5.  
A:Reference number: Z14470  
A:Accession: T01953  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-190 <GEI>  
A:Cross-references: UNIPROT:082604; UNIPARC:UPI00000A8F7D; EMBL:AF096371; NID:g3695386;  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4  
A:Introns: 36/2; 69/3  
A:Note: T2L5.6  
C:Superfamily: Arabidopsis thaliana hypothetical protein T2L5.6

Query Match 5.8%; Score 162.5; DB 2; Length 190;  
Best Local Similarity 28.6%; Pred. No. 3.3e-05;  
Matches 54; Conservative 34; Mismatches 62; Indels 39; Gaps 10;

Qy 375 QVFFGAGNYHLVD-ENFPELPDYVLSLFLKLVGPKVLMRSRVKGPDRSKLRYVLHCTNV 433  
Db 12 QSLIG-GNYGLLNTNFTPNFDYYSALIWRQLMKRKALFTTFSG--TKIRSYTHCA-- 66

Qy 434 YHPRYREGDLTYLVNLHNV-----TKHLK-----LPPMFESRP----- 467  
Db 67 ---RQSKG-ITVLLMNLDNTTVAKVELNNSFLRTHKMKSYKRASSQVFGGNGVIQ 122

Qy 468 VDKYLLKPFSGDG-LLSKSVQLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPAFSYGPFV 526  
Db 123 REEYHLT--AKDGNLHSQTMLLNGNALVNSMGDLPPIEPHINHSTPITAPYSIVFVH 180

Qy 527 IRNAKIAAC 535  
Db 181 MRNVVVPAC 189

RESULT 4  
T45668  
hypothetical protein F14P22.70 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 09-Jul-2004  
C:Accession: T45668  
R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Lemcke, K.;  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: Z23011  
A:Accession: T45668  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-575 <DAN>  
A:Cross-references: UNIPROT:Q9M2G8; UNIPARC:UPI000009B66E; EMBL:AL137082  
A:Experimental source: cultivar Columbia; BAC clone F14P22  
C:Genetics:  
A:Map position: 3  
A:Introns: 2/1; 126/2; 164/3; 186/3; 216/2; 245/3; 325/3; 359/3  
A:Note: F14P22.70  
C:Superfamily: Arabidopsis thaliana hypothetical protein F8J2.40

Query Match 3.9%; Score 109; DB 2; Length 575;  
Best Local Similarity 22.9%; Pred. No. 2.4;  
Matches 102; Conservative 57; Mismatches 140; Indels 146; Gaps 25;

Qy 38 TKRLFSQSPSLFITIDASLATDP-----RFLTFLGSPRLRLARGLSPAYLRFEG 89  
Db 3 TKTLRSI--SFNDIDSDDSTITESPEARLYNRSLSGKGTQRSR-----ISP-----GD 50

Qy 90 TKTDFLI--FDPNKEPTSEERSYWSQDNNDICGSERVASDVLKRLQNEWFPQELLRLRE 147  
Db 51 VENNFSIKPLTPMKDDRYKMIWK-----PVCNEH-AKEFLALLGDGTGYQAALKQK 104



```

Qy 148 QYORFKNSTYSRSSVD-----MLYSFAKCSRLDLIFGLNALLRTDPLRNSSNA 197
Db 105 VY-RSFRP---RRRLADCAVVVEQRWWKVLDFAELKRSSIISFFETEKQETAIVSRMSRART 160
Qy 198 QLLLYNCSSKGYNIS-----WELGNEPN-----SFWKKAQISIDGLQLGED 238
Db 161 R-----AAKVGKGLSKDEKARKALQHLEAIDPRHRYGHNLQFYHHAWLHCD----- 208
Qy 239 FVELHKLLQSAFONAKLYGPDIGQ-----PRGK-----TVKLRSFLKAGGEVIDS 285
Db 209 -----SKQPF-----FYWLIDIGOGKELNHRCPSRKLYQQSIKYLGPTEREAIVEIIE 256
Qy 286 LTMHHYYLNGRVATKEDFISSDVLT-----FILSVQIKLVKTVMTPGKKVWLIG 335
Db 257 -----DGKLMYKO--SGVILDTKEGPPDAKWIFVLVSVKILYGM-----KKGNF 300
Qy 336 ETSSAYGGAPLNSNTFAAGFMWLKGLSAQLGIEVVMRVQFFGAGNYHLVDENFEPLP 395
Db 301 QHSFLAGGA-----TLSAGRIIVD-----DGVLKAVMPHSHGYLPTEENFOA--- 343
Qy 396 DYWLSILLFKLYCPKPKVMSRVKGPD 420
Db 344 --FMSFLRENNVD---LANVKNQND 363

RESULT 5
A48954
mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum
N;Alternate names: beta-mannanase
C;Species: Caldocellum saccharolyticum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A48954; BA3745
R;Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.
Appl. Environ. Microbiol. 58, 3864-3867, 1992
A;Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain
A;Reference number: A48954; MUID:93119139; PMID:1476429
A;Accession: A48954
A>Status: preliminary
A:Molecule type: nucleic acid
A;Residues: 1-1331 <GB>
A;Cross-references: UNIPROT:P22533; UNIPARC:UIP000012EB88; GB:L01257; NID:g144290; PIDN:
A;Note: sequence extracted from NCBI backbone (NCBIN:121576, NCBIPI:121577)
R;Luthi, E.; Bhana Jasmat, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.
Appl. Environ. Microbiol. 57, 694-700, 1991
A;Title: Cloning, sequence analysis, and expression in Escherichia coli of a gene coding
A;Reference number: A43745; MUID:91247819; PMID:2039230
A;Accession: BA3745
A>Status: preliminary
A:Molecule type: DNA
A;Residues: 1-337,'PPRQHOHRQ' <LWE>
A;Cross-references: UNIPARC:UIP000016EB78; EMBL:M36063; NID:g144292; PIDN:AAA72861.1; PT
A;Note: the authors translated the codon CAC for residue 262 as Glu
A;Note: this sequence has been revised in reference A48954
C;Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 3.9%; Score 109; DB 2; Length 1331;
Best Local Similarity 18.3%; Pred. No. 8.4;
Matches 104; Conservative 75; Mismatches 196; Indels 192; Gaps 23;

Qy 21 TPAGTA-PFKQVVDLFYTKRLFSQVSPSLITIDASLATDPRLFTFLGPSRLRAL---- 76
Db 758 TPTVTATPTPTPTPIPTVTPPLPTISPSVVEITINTNAGRTOI-----SPYIYGAND 811
Qy 77 ARGLSPAYLRFGCTKTDFLFDPNKEPTSEERSYWQSDNNDICGS-----ERVSADV 129
Db 812 IEGVVHSAARLGNRLTGYNWENN---FSNAGNDWYHSSDDYLCWSMGCISEDAKVPAV 868
Qy 130 LRKLQWEPFQELLRLREOYREFKNSTYSRSSVDMLYSPAOKS----- 173
Db 869 VSKF-----HEYSLKNNAYSAVTLQMGAYSKDNVGTYSNETAPSNRWA 913
Qy 174 ----RLDLIFGLNALLRTDPLRNSSNAQLLNLYNCSSKGYNIS-----WELGNEPN 221

```

Db 914 EVFKKDAPLSLN-----PDLNDFVYMDDEFINYLINK-YGMASPTGIGKIYILDNDPD- 966  
 Qy 222 FWKKAQISIDG-----LQLGEDFVELHKLQKSAFQNAKLYG-----PDI 261  
 Db 967 LWASTHPRIHPNKVCKTELIEKSEVELAKVI-KTLDPSEAVFGVASYGFMGYSIQLQADPW 1025  
 Qy 262 GQPRGKTVKLLRSFLKA-----GGEVTDISTWHHY----- 291  
 Db 1026 NOVKGHRWFIPTWLEQMKKASDFGRLLDLVDLDLHWYPEARGNIRVCFDGENDTSKVEY 1085  
 Qy 292 -----YLNGRVATKEDFLSSDVLDTFTLSVQKILKVTKEMTPGKKVW 333  
 Db 1086 VIARMQAPRTLMDPTYKTSVKGQITAGENSINQWFSYDILPIIPNVKADIEKYYPGKJLA 1145  
 Qy 334 LGETSSAYGGCAPLLNSTFAAGFMWLDKGLSLAQLGIEVVMR-----QVF--- 378  
 Db 1146 ISEFD--YGG---RNHISGGIALADVLGIFGKYGVNFAARWGDSSGAAAAVNIYLYN 1198  
 Qy 379 -----FGAGNYHLVDENFELPDYVLSLLFKLVGPKVLMRSVKGPDRKLRVYLHCTN 432  
 Db 1199 DGKSKYGTNTVSANTSVDENMPVY-----ASINGQDDSELHIIILNRN 1242  
 Qy 433 VYHPRYREGDLTYV-LNLHNVTKHLK 458  
 Db 1243 Y-----DQKLQVKINITSTPKYTK 1261

RESULT 6  
 E71086  
 hypothetical protein PH0954 - Pyrococcus horikoshii  
 C:Species: Pyrococcus horikoshii  
 C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 12-Jul-2004  
 C:Accession: E71086  
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, N.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.  
 DNA Res. 5, 55-76, 1998  
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic Pyrococcus horikoshii strain OT3  
 A:Reference number: A71000; MUID:98344137; PMID:9679194  
 A:Accession: E71086  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-4436 -KAW>  
 A:Cross-references: UNIPROT:O58659; UNIPARC:UPI00001100E3; GB:AF000004; NID:932  
 A:Experimental source: strain OT3  
 A:Note: this accession replaces an interim accession for a sequence replaced by C:Genetics:  
 A:Gene: PH0954

Query Match 3.9%; Score 108.5; DB 2; Length 4436;  
 Best Local Similarity 19.6%; Pred. No. 56;  
 Matches 101; Conservative 73; Mismatches 201; Indels 139; Gaps 22;

Qy 77 ARGLSPAYLRFGGTKDTFLFDNKEPTSEERSYV--QSODNNDICGSESVSADVLRLKQL 134  
 Db 1018 SRDVTILRYLSPSSGKHNLISVLDPYNR-----WIEENEENNLFTLSFGKPDLVKEG 1069  
 Qy 135 MEW-----PFQELLRLREYQREFKNST---YSRSSVDMLYSF 169  
 Db 1070 ITWAPNYFTSGENVLFYIVKNLQGPFLKSFYVR---AZIWNTRKIYSTNAYPRNWSF 1125  
 Qy 170 AKCSRLDLIFGLNALLRTPDLRNSSNA-----QLLLNYCSS-----KGYNISWELG 216  
 Db 1126 GKG-----ETKEFNWRWYNAKPGNLTVKTIWVDYDYNISPEGNSNNNEFSAFGL 1172  
 Qy 217 NEPNRSFWKKAQISIDGLQGEDFVELHKLQKSAFQNAKLYGDPDIPGRCKTVKLLRSFL 276  
 Db 1173 NVGTPDFLENLSVEDLAYGK-FVRIN-----ATVKNLGDISI 1208  
 Qy 277 KAGEVIDSUTWHHYLYNNGRVATKEDFLSSDVLDTFTLSVQKILKVTKEMTPGKKVWLGE 336  
 Db 1209 YRPITVLFNVSGERYY-RTVYGIKIKENESKSVTLPTWYVDRVGEY-FRVKEVDPGNRIIVEGN 1266

```
QY 337 TSS-----AYGGAPLLSNTFAAGMWMLDKGLLSAQLGIEVVMR-----QVFFGAGNYHLV 387
Db 1267 ESNIIIRYYVESPEL---MLSGIEWLEEEVVRGYLAYKVNVTNTGGDVIRGYVMFV 1323
QY 388 DENFEPDPYWLSSLFKKLWGPVKLSRVKGPDRSKLRVYLHCTNVVHPHYRREGDLTYV 447
Db 1324 DG--EPKSSWINKLLHGETAERTLRWRFSSGGKKEVRIIVVDQD-YIPESNEDNNAI-- 1378
QY 448 LMLHNVTKHLKP-----PPMFS-----RPVDKYLLKLPFGSGGLLS 483
Db 1379 --VENVT--IVLPDIEVLSNIPSMHANSYFKVNATIKNSGGQDVKRIFVYVSLYQDGKLL 1434
QY 484 KSVQLNGQTLKWYDEOTLPALTEKPLPAGSSLSV 517
Db 1435 GSAPVYSLASGEVKEVT---LTIRPYGNSTFKV 1465

RESULT 7
S01042
aerobactin receptor precursor iutA [validated] - Escherichia coli plasmid ColV-K30
N:Alternate names: cloacin receptor
C:Species: Escherichia coli
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 09-Jul-2004
C:Accession: S01042
R:Krone, W.J.A.; Stegehuis, F.; Koningstein, G.; van Doorn, C.; Roosendaal, B.; de Graaf
FEMS Microbiol. Lett. 26, 153-161, 1985
A:Title: Characterization of the pColV-K30 encoded cloacin DF13/aerobactin outer membran
e sequence and primary structure.
A:Reference number: S01042
A:Accession: S01042
A:Molecule type: DNA
A:Residues: 1-725 <KRO>
A:Cross-references: UNIPROT:P14542; UNIPARC:UPI000017AA31; EMBL:X05874
C:Genetics:
A:Gene: iutA
A:Genome: plasmid ColV-K30
C:Function:
A:Description: functions as outer membrane receptor for ferric aerobactin [validated, MU
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-725/Product: cloacin receptor #status predicted <MAT>

Query Match 3.9%; Score 108; DB 2; Length 725;
Best Local Similarity 20.4%; Pred. No. 4;
Matches 79; Conservative 53; Mismatches 154; Indels 102; Gaps 14;
QY 80 LSPAYLRFGGTKDFLIFDPNKEPT--SEERSYWSQDNNDCGSSERSVADVLKQIM-- 135
Db 196 LSVAYQKFGW-----FDGNGDATLLDNTQTGLQYSDRLDIMGTFGLNIDESRQLQILIT 249
QY 136 -----EWPFBELLILREQYQ-----RE 152
Db 250 QYKSGDDDYGLNLGKGFSAIRGTSTPPVSNGLNSDRIPGTERHLILSLQYSDSAFLGQE 309
QY 153 FRNSTYSRSSVDMLYSFAKCSRDLDFGLNALLRTPDLRWNSNAQLLLNYCSKGYNIS 212
Db 310 LVGVVYRDESILRFYFPPTWNANKQVTAFFSSQQDTD---QYGMKLTLSNKPMDGWOIT 365
QY 213 WELGNEPNSFWKKAQISIDGLQGEDFVELHKLQSAFONAKLYGDPDIOQPGKTVKLL 272
Db 366 WGLDADHERFTS-----NOMFFDLAQSASGGLNKKIY--TTGRYPSYDITNL 412
QY 273 RFLKAGGEVIDSLTWHHYLNGRVATKEDFLSSDVLDTFLSVQKILKVTKEMTPGKV 332
Db 413 AAFLOSGYDINLFT-----LNGGVRYQ-----YTENKIDDFIGVAAQ-----RQIGAK-- 457
QY 333 WLGETSSAYCGGAPLLSNTFAAGFM-----WLDKGLSLAQLGIEVVMRQVFFGAGN 383
Db 458 --ATSADAFWLSRLRHLFLNAGLLMHIITPQAM-----LNFSGVELPDPGKYVGRGI 510
QY 384 YHLVDENFEPLPDYWLSSLFKKLWGPVK 411
Db 511 YGAANVGNHLP LTKS-VNVS DSKLEGVKV 537
```

## RESULT 8

```
A31842
endo-1,4-beta-xylanase (EC 3.2.1.8) Z precursor - Clostridium thermocellum
N:Alternate names: xylanase Z
C:Species: Clostridium thermocellum
C:Date: 31-Mar-1990 #sequence_change 11-Apr-1997 #text_change 09-Jul-2004
C:Accession: A31842
J:Grepinet, O.; Chebrou, M.C.; Beguin, P.
R. Bacteriol. 170, 4582-4588, 1988
A:Title: Nucleotide sequence and deletion analysis of the xylanase gene (xynZ) of Clostri
A:Reference number: A31842; MUID:89008072; PMID:3139632
A:Accession: A31842
A:Molecule type: DNA
A:Residues: 1-837 <GRE>
A:Cross-references: UNIPROT:P10478; UNIPARC:UPI000013909C; GB:M22624; NID:gl44931; PIDN:f
C:Genetics:
A:Gene: xynZ
C:Function:
A:Description: catalyzes the hydrolysis of 1,4-beta-xylosidic linkages in xylans
A:Pathway: xylan degradation
C:Superfamily: Clostridium endo-1,4-beta-xylanase Z; Clostridium cellulase repeat homolog
C:Keywords: duplication; extracellular protein; glycosidase; heat-stable protein; hydroli
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-837/Product: endo-1,4-beta-xylanase #status predicted <MAT>
F:326-419/Domain: Clostridium xylanase A repeat homology <SCXA>
F:430-453/Domain: Clostridium cellulase repeat homology <CCRI>
F:464-487/Domain: Clostridium cellulase repeat homology <CCR2>
F:548-834/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>
F:645,754/Active site: Glu #status predicted

Query Match 3.8%; Score 107; DB 1; Length 837;
Best Local Similarity 19.4%; Pred. No. 6;
Matches 86; Conservative 64; Mismatches 135; Indels 158; Gaps 23;
QY 14 LRALTQGTAGTAPTAKDVLDLEFYYTKRLFQSVSFSLSI-TIDASLATDPRFLPLGSPR 72
Db 484 LRITTEPFGGDVQTFN-----PSVTPTQTPIPTISGNALRD----- 520
QY 73 LRALARGLSPAYLRFGGTKDFLIFDPNKEPTSEERSYWSQDNNDCGSSERSVADVLK 132
Db 521 -YAEARGIKI-----GTCVNYPPYN-NSDPT--YNSILQREFSMVVCENE-MKFEDALQP 569
QY 133 LQMEWPFQELLRLREQYQREFKFNSTYSRSVDMLYSFAKSRDLDFGLNALLRTPDLRW 192
Db 570 RQNVDFP-----SKGDQLLAFAERNMQO-----MRGHTLIW 600
QY 193 NSSNAQLLIN-----YCSSKGYNISWELGNE-----PNSFWK 224
Db 601 HNQNPSWLITNGNWRNDSLLAVMKNHITVTMTHYKGIKIVEMDVANECMDSDSGNGLRSSIWR 660
QY 225 KAQISIDGLQGEDFVELHKLQSAFONAKLYG-----PDIGQPRGKTVKLLRSFLKAG 279
Db 661 NV-----IGQDYLDYAFRYAREADPDALLFYNDYNIEDLGPKSNVAFNMKS-MKER 711
QY 280 GEVIDSLTWHHYLNGRVATKEDFLSSDVLDT-----FILSVQKI-LKVTKEMTPG 329
Db 712 GVPIDGVGFQCHPING---MSPEYLAS-IDQNKYAEIGVIVSFTIDIRIPOSENPA 766
QY 330 -----KKVWLGETSSAYCGGAPLLSNTFAAGFM--LDKUGLSAQLGIEVVM 374
Db 767 TAFQVQANNYKELMKIKLANPN-----CNTFV---MWGFTDKY-----TWI 804
QY 375 RQVFFGAGNYHLVDENFEPLPDY 397
Db 805 PGTFFPGYGNPLIYDSNYPNKPAY 827
```

## RESULT 9

```
T50113
3-dehydroquininate synthase (EC 4.2.3.4) - fission yeast (Schizosaccharomyces pombe)
N:Contains: 3-dehydroquininate dehydratase; 3-dehydroquininate synthase (EC 4.6.1.3)
C:Species: Schizosaccharomyces pombe
```

C;Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jul-2004  
C;Accession: T50113  
R;Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, February 2000  
A;Reference number: Z25039  
A;Accession: T50113  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1573 <SEE>  
A;Cross-references: UNIPROT:Q9P7R0; UNIPARC:UPI0000125F22; EMBL:AL157734; PIDN: CAB75770.  
A;Experimental source: strain 972h(-); cosmid c1834  
C;Genetics:  
A;Gene: SPDB:SPAC1834.02  
A;Map position: 1  
A;Superfamily: pentafunctional Arom protein; 3-dehydroquininate dehydratase homology; 3-dehydroquininate dehydratase homology  
C;Keywords: shikimate kinase homology  
F;407-635/Domain: 3-phosphohikimate 1-carboxyvinyltransferase homology <PSK>  
F;1035-1279/Domain: 3-dehydroquininate dehydratase homology <DQD>  
  
Query Match 3.8%; Score 106.5; DB 2; Length 1573;  
Best Local Similarity 20.3%; Pred. No. 17;  
Matches 115; Conservative 85; Mismatches 190; Indels 177; Gaps 30;  
  
QY 19 QGTGAGTAPTKVDVLEFYTKLQSPSPSFLSITIDASLAT-----DP-RFLTFLGSPR 72  
DB 699 QGPPKGLKPLESIDME-----TWTDAPLTASVVAACNVSEGDVPVTRITGIANQR 750  
  
QY 73 LR-----ALARGSPAYLFGGKTDFLFDPN-KEPTSEERSYQSQDN----- 116  
DB 751 VKECNRITAMVHELAKGVTGTELEDGIYIFGKYKELKPEEGIYYDDHRIAMSVSL 810  
  
QY 117 NDICGSEVSADVLRLKQMEWPFQELLRLREYQREFKNSYSSVDMLYSFACSRDL 176  
DB 811 SLICPSRTLIID-KACVEKTPYWM-VDLHQSGFGVKLTGAT-SVASDPLKGSISKNSAII 867  
  
QY 177 LIFGLNALLRTP-----DLRWNSSNAQLLLNCSSKGYNSWELGNENPSF 222  
DB 868 LI-GMRGAGKTTICKITAKQLNFKFLDL-----DELLEDYLEMPIAEVIFRMG----- 914  
  
QY 223 WKKAQISIDGLQGEDFVELHKLQSAFQNAKLYGPDIGOPRKTKV-----LLRSFL 276  
DB 915 W-----DAFLEE-----HKVLRKFTIHEPEGY-----VAASGGVVIEMDESRLNLFV 959  
  
QY 277 KAGGEVIDSLTHYLYNGRVATKEDFLSSDVLDTFILSVQKILKVKEMTPGKKWVLGE 336  
DB 960 KEGIVL-----HVHNRN--LEHKVLSYSEDQTPYTKQDSIDDDYKR-----RHVVYRE 1007  
  
QY 337 TSSAYGGAPLLSNTFAAGFMWLDKGLSALQGLIEVVMRQVFFGAGNYHLVDENFEPLPD 396  
DB 1008 CRSHY-FISPVLSN-----QVIDEKIQ----- 1028  
  
QY 397 YWLSLFLKLVGPKVLMRVKGPDRSKRLRVYLHCTNVTYHPRYREGDLTYLVNLHNTKH 456  
DB 1029 YMSRFLDVTGSSQVLQKFKTKRSTF-----LTNYPRIEDALPTL-----RD 1073  
  
QY 457 LKLPMPFSPVDKYLKPKFGSDGLSKSVQLNQTLKMWDEQ-----TLPAL-TEK 507  
DB 1074 VTGCDIAEVRVD-YLKDPPSSNGISS-----LDFVAEQISLLRCSTTLPIITIR 1123  
  
QY 508 PLPAG-----SSLSVPAFSYG 523  
DB 1124 TISQGLFPNDKEEAKELMSARVY 1150  
  
RESULT 10  
S75363  
hypotheetical protein sll1913 - Synechocystis sp. (strain PCC 6803)  
C;Species: Synechocystis sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C;Accession: S75363  
R;KaneKo, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.

O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis s.  
A;Reference number: S74322; MUID:97061201; PMID:8905231  
A;Accession: S75363  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-599 <KAN>  
A;Cross-references: UNIPROT:P73250; UNIPARC:UPI00000C0CB6; EMBL:D90904; GB:AB001339; NID: A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
  
Query Match 3.8%; Score 106; DB 2; Length 599;  
Best Local Similarity 19.6%; Pred. No. 4.3; Mismatches 220; Indels 168; Gaps 31;  
Matches 114; Conservative 79;  
  
QY 6 LLLW---LWGR-LRALTQGTAGTAPTKDWDVLEFYTKLQSPSPSFLSITIDASLATD 61  
DB 64 LAAMQTGLWGEFLRTLTVQPGG-----LDHSWVSLF-TFDPRL-GSKIFADWAAE 112  
  
QY 62 PRFLTFLGSP-RLRALARGLSPAYLRF-----GGTKT-DPLIFDPNKEPTSEER 108  
DB 113 LPNLTWTIANQCPMEVVKNGDRLVGVRFPDYETRARVILDTGTELGLDALLG-----DIGHR 167  
  
QY 109 SYMQSQDNNDICGSEVSADVLRLKQMEWPFQELRLREYQREFKNSYSSVDM 165  
DB 168 WGMWQDKRFD-----EPSCPIAPNEMTEBYPIQSPTWVFLRLKRTNNQ---TNIAEPTIDV 221  
  
QY 166 LYSFAKSRDLIFGLNALLRTPDLRWNSSNAQLLLNCSSKG---YNISWELGNENPSFW 223  
DB 222 AQDPTH-----TWQNYGEKDFLYGQLPGEHYMINWPIAG--NDYG 260  
  
QY 224 KKAQISIDGLQGEDFVELHKLQSAFQNAKLYGPDIGOPRKTKVLMRSLFKAGGEVI 283  
DB 261 KDLNRLLGGEKEKQTYL-----KEAQYSYAYVYLQKHHSNLELATGIPQTGDIS 313  
  
QY 284 DSLTWHHYLYNGR-----VATKEDFLSSDVLDTFILSVQKILKV-----T 323  
DB 314 TAFALHPYRESRRLKGOQVITERDILPQGVASLPIYNQKVTSGVGNVANDHHYPGYE 373  
  
QY 324 KEMTPGKKWVLGE-TSSAYGGAPLLSNTFFAAGFMWLDK----- 361  
DB 374 FPLTPKSLIWGRWGTGPTTTPFPALLSTARGYLPCERKNISVSHWANGSTRLOPLVMNT 433  
  
QY 362 ---LGLSAQLGIEVVMRQVFFGAGNYHLVDENFEPLDYWLSLFLKLVGPKVLMRSLVK 418  
DB 434 QQAVGMIALSVE-----KNCDP-QDLDVDVDQALIG----- 465  
  
QY 419 PDRSKRLRVYLHCTNVTY--HPRYREGDLTYLVNLHNTKHKLKPLPPMFSRPVDKYLK-P 475  
DB 466 -DRRAPAAVIFLNFNLVDPDHPERWQW--QYVLD--NPDQY----PPSGHCPVDVNLQALP 516  
  
QY 476 FGSD-----GLLSKSVQLNGOTLKNVDEOT---LPALTEKP 508  
DB 517 LSKSQSVYTGELQKS---EHQTYQLICQSQSGKILKVITERP 554  
  
RESULT 11  
RNZ82  
genome polyprotein - human respiratory syncytial virus (strain A2)  
N;Alternate names: polymerase L protein  
N;Contains: RNA-directed RNA polymerase (EC 2.7.7.48)  
C;Species: human respiratory syncytial virus  
A;Note: host Homo sapiens (man)  
C;Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 09-Jul-2004  
C;Accession: A40317; A28319; PS0048  
R;Stec, D.S.; Hill III, M.G.; Collins, P.L.  
Virology 183, 273-287, 1991  
A;Title: Sequence analysis of the polymerase L gene of human respiratory syncytial virus  
A;Reference number: A40317; MUID:9127486; PMID:2053282  
A;Accession: A40317  
A;Molecule type: mRNA  
A;Residues: 1-2165 <STE>

A; Cross-references: UNIPROT:P28887; UNIPARC:UPI0000134AEA; GB:M75730; NID:g333955; PIDN:R; Collins, P.L.; Olmsstead, R.A.; Spriggs, M.K.; Johnson, P.R.; Buckler-White, A.J. Proc. Natl. Acad. Sci. U.S.A. 84, 5134-5138, 1987

A; Title: Gene overlap and site-specific attenuation of the viral polymerase

A; Reference number: A28319; MUID:87260943; PMID:2440043

A; Accession: A28319

A; Status: preliminary; not compared with conceptual translation

A; Molecule type: DNA

A; Residues: 1-81 <COL>

A; Cross-references: UNIPARC:UPI0000134AE9; GB:M17245; NID:g333953; PIDN:AAA47417.1; PID:J. Gen. Virol. 69, 2901-2906, 1988

A; Title: The A and B subgroups of human respiratory syncytial virus: comparison of internal

A; Reference number: P30048; MUID:89036169; PMID:3183631

A; Accession: P30048

A; Molecule type: mRNA

A; Residues: 1-18 <JOH>

A; Cross-references: UNIPARC:UPI000017272; GB:D00397; NID:g222551; PID:g2160375

A; Experimental source: strain 18537

A; Note: this strain belongs to subgroup B

C; Genetics:

A; Gene: L

C; Superfamily: parainfluenza virus RNA-directed RNA polymerase

C; Keywords: ATP; nucleotidyltransferase

Query Match 3.8%; Score 106; DB 1; Length 2165;  
Best Local Similarity 18.8%; Pred. No. 30;  
Matches 116; Conservative 95; Mismatches 216; Indels 190; Gaps 29;

Qy 31 VVDLEFYTKRLFSQVSPFLSTITIDASLAT-DPRFLT-----FLGSPRLRALARGLSP 82  
Db 1199 ITELSKYVRERSWSLS-NIVGVTPSMTYMDIKYTTSTISSGIIIEKYVNSLTRG--- 1254

Qy 83 AYLREGGTTDFLI FDPNKEPTSEERSYWSQDNNDICGSEVSADVLKLOMEWPFQEL 142  
Db 1255 ---ERGPYK-----PWGSSSTQKEKTPVYARQVLTKQRQDIDLAKL--DMVYAS- 1301

Qy 143 LLLREOYQREFK-----NSTYSRS-----SYDMLYSFAKCSR-----LDL 177  
Db 1302 IDNKDFEWEELSIGTLGTYERAKKLFPQVLSVNYLHRLTVSSRCPBPAPAYRTNY 1361

Qy 178 IFGLNALLRTPDLRWNSSNAQLLNYCSSKGYNISWELGNPN-----SPFKKAQIS 230  
Db 1362 HFDTSPIKRLTEKYGDEIDIVFQNCISFGLSLSVVEQFTNVCNRIILIPKLINEHL 1421

Qy 231 DGLQGEDFVELKLLKSAFQAKLYGPD-IGQPRGKTVKLLRSFLKAGGEVDSLTHW 289  
Db 1422 MKPPIFTGDVDIHKL--KQVIQKHMFLPDKISLTQYVELFSLNKLKSGSHVANSNLIA 1479

Qy 290 H----YYLNGRVATKREDFLSDVLDFTILSVQKILKYTKEMTPGKKVWLGETSSAYGGA 345  
Db 1480 HKISDYFNNTYI-----LSTNLAGHWILLIQ-LMKDSKGIF--EKDW-----GEG 1521

Qy 346 PLLSNTFAAGFMWLDKLGLSAQLGIEVVMRQVFPFGAGNYHLV-----DENFE 392  
Db 1522 YITDHMF-----INLKVFNFAYKTYLLCFHKGYGKAKLECDMNTS 1561

Qy 393 PL-----PDYWLSL-----LFKKLVGPKVLMRSRVKGPDRSKLRVYLHCTNV--- 433  
Db 1562 DLLCVLELIDSSYKWSMSKVFLPEQVKIYILSDQASLHRVKRGCHSFKL-WFLKRLNVAEF 1620

Qy 434 -----YHPRREGDIT-----LYVLNL----- 450  
Db 1621 TVCPMVVNIDYHTPMKAILTYIDLRVMGLINIDRIHKNKHFNDEFTYTNLFYNTNF 1680

Qy 451 ----HNVTKHLKLPDPMFSRPVDK-YLLKPFSGDLILSKSVQLNG-OTL-----KMYDE 498  
Db 1681 SDNTHLLTKGIRIANSELENNYKLYHPTETLENILANPKSNCKKTLNDYCIKQVDS 1740

Qy 499 QTLPALTEKPLPAGSSL 515  
Db 1741 IMLPLLSNKKLIKSSAM 1757

[illegible]

A:Residues: 668-690, 'M', 692-753 -MAR>  
A:Cross-references: UNIPARC:UPI0000176348; EMBL:X51541  
R:Sotttrup-Jensen, L.; Sand, O.; Kristensen, L.; Fey, G.H.  
J. Biol. Chem. 264, 15781-15789, 1989  
A:Title: The alpha-macroglubulin bait region. Sequence diversity and localization of cleavage sites in the alpha-macroglubulin bait region.  
A:Reference number: A34230; MUID:89380162; PMID:2476433  
A:Accession: A34230  
A:Molecule type: protein  
A:Residues: 670-752, 'Q', 754-759 <SOT>  
A:Cross-references: UNIPARC:UPI0000176349  
C:Genetics:  
A:Gene: GDB:P2P  
A:Cross-references: GDB:120330; OMIM:176420  
A:Map position: 12p13-12p12.2  
A:Introns: 73/3; 116/2; 146/3; 169/3; 703/1; 753/2  
A:Note: the list of introns may be incomplete  
C:Superfamily: alpha-2-macroglobulin  
F:685-735/Region: bait region

Query Match 3.7%; Score 104.5; DB 2; Length 1482;  
Best Local Similarity 20.8%; Pred. No. 22;  
Matches 113; Conservative 72; Mismatches 188; Indels 169; Gaps 29;

Qy 9 WLWGLRALTQGT-----PAGTAPTQDVVDLEFYTKRLQFSVSPSLSTI 54  
Db 745 WIWELVAVNSSGVAEVGTVPTDITWKAQAFCLSEDAGLGISSTASURAFQPPFVELTM 804  
Qy 55 DASLATDRPF-----LTFILG-----SPRLRALARGSPAYLRFQGTXTDFLIPDNKEP 103  
Db 805 PYSVIRGEVFTLKATVNLVLPKICIRSVQLKA-----SPAFASQNTK----- 847  
Qy 104 TSEERSYMQSDNDICGSESVADVLRLQMEWPFQELLRLRLQYQREFKNSTYSRSV 163  
Db 848 --GEESYC-----ICGSEROT-----LSMTVT-----PKTLGNVNES-VSA 880  
Qy 164 DMLYFAKCSRLDLIFGLNALLRTPDLRWNSSNAQLL-----NYCSSKGVN 210  
Db 891 EAMQSLELCG-----NEVEVPEIKRKDTVIKTLVVEAGIEQEKTFSSMTCASGAN 932  
Qy 211 ISWELGNE-PNSFWK--QAQISIDGLQGEDFVELHKLQK---SAFQNAKLYGPDIG- 262  
Db 933 VSEQLSLKLPNSVVKESARASFVLDGLILSAMQNIQNLLQMPYCGGQNVLPAPNIYV 992  
Qy 263 -QPRGKTVKLRSF-LKAGEVID----SLTMHHYLLNGRVAT-----KEDFLSSD 307  
Db 993 LNYLNETQQLTQEIKAQAVGLITGYORQLNYKHQ--DGSYSTFGERYGRNQNTWLTAF 1050  
Qy 308 VLDTFILSVQKIL----KYTKEMTPGKKWLGE---TSSAYCGGAPILSNTPFAAGFMWLD 360  
Db 1051 VLKTFQAQARSYFIDEAHITQSUT-----WLSQMOKDNGCFRSGSLNNAIKGVE--D 1103  
Qy 361 KLGLSAQLGIEV-----VMRQVFF-----GAGNYHLVDENFEPULPDYWL 399  
Db 1104 EATLSAVTIALLEIPLVPTNPIRNALFCLSEAMNVAKETHGSHVYT---KALLAYAF 1160  
Qy 400 SLLFFKKLVGPKVLMRSVK-----GPDRSKLRV-YLHCTNVYHPRYREGDLTLY 446  
Db 1161 SLLGKQNQREILNSLDKEAVKEDNLVHWRPQRPKAPVGHLYQTQA---PSAEVEMTSY 1217  
Qy 447 VL 448  
Db 1218 VL 1219

RESULT 14  
S61166  
Probable membrane protein YDR371w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein D9481.7  
C:Species: Saccharomyces cerevisiae  
C:Date: 23-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 09-Jul-2004  
C:Accession: S61166  
R: Ding, H.  
Submitted to the EMBL Data Library, June 1995

A;Description: The sequence of *S. cerevisiae* cosmid 9481.  
A;Reference number: S61159  
A;Accession: S61166  
A;Molecule type: DNA  
A;Residues: 1-511 <DIN>  
A;Cross-references: UNIPROT:Q06350; UNIPARC:UPI000006A221; EMBL:U28373; NID:g849184; PID:  
A;Experimental source: strain S288C (AB972)  
C;Genetics:  
A;Gene: MIP5:YDR371W  
A;Cross-references: SGD:S0002779  
A;Map position: 4R  
C;Superfamily: Serriatia marcescens chitinase  
C;Keywords: transmembrane protein  
F;18-34/Domain: transmembrane #status predicted <TMM>

Query Match 3.7%; Score 104; DB 2; Length 511;  
Best Local Similarity 19.8%; Pred. No. 4.9;  
Matches 72; Conservative 52; Mismatches 132; Indels 108; Gaps 15;

Qy 200 LLYCSCSGYNI SWELGNEPNSFWKQAIISDGL-----QLGEDFVELHKL 246  
Db 165 LKNTCSKKPKVIMSIGWSSENFKIIIKDKLLQNFVDSVETMFRIGFDGIDL---- 220  
Qy 247 QKSAFONAKLYGPDIGOPRG--KTVKLLRSFLKA-GGEVIDSLTWHHYVYNGKRVATKEDF 303  
Db 221 -----DWEPPGNESRPGYLVLRMLRLKLSLESQIFGKRTEDHFOLSTAAPAFKOK 274  
Qy 304 LSSDLVDTFILSVQIKLVTKEMTPGKKVWLGETSSAYGGGAPLLSNT-----FAAGPWW 358  
Db 275 L-----FVLPIEIDQYVDYNNMTVDYIGSWSETTGYHSNLFSETELNGNFAMHYW- 326  
Qy 359 LDKLGLSAQIGIEVVMQVFFGAGNYHLVDENPEPLPD--YWLSSLFFKLVCVKVLMRSRV 416  
Db 327 IDRFGVNSR---KLVLGMAAYGR-SPHIKONKFEFPNQNTVLINKIFKGVGKPTKEIDKA 382  
Qy 417 KGPD-----RSLRVLYLH----- 429  
Db 383 DGREGIWPYKNLPKIGTIEQVDPKVGSAYCFDEKNSIFISYDNTKSVTKAEYVTHNLLG 442  
Qy 430 -----CTNVYHPRYREGDITLVNLHNVYTKHLKL--PPPMFSR-PVDKYLKLPFGSD 479  
Db 443 GGFWESCGEAYANESRS-----LINAFNEGLHFNVSCKPSIFQDVRVKYKYLKNGYGDG 496  
Qy 480 GLLS 483  
Db 497 GFLS 500

RESULT 15  
T44483  
receptor-like protein iutA [imported] - Shigella flexneri  
C;Species: Shigella flexneri  
C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C;Accession: T44483  
R;Moss, J.B.; Cardozo, T.J.; Zychlinsky, A.; Groisman, E.A.  
Mol. Microbiol. 33, 74-83, 1999  
A;Title: The selC-associated SHR-2 pathogenicity island of Shigella flexneri.  
A;Reference number: Z22779; MUID:99340540; PMID:10411725  
A;Accession: T44483  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-732 <MOS>  
A;Cross-references: UNIPROT:Q9XCH0; UNIPARC:UPI00000BE41C; EMBL:AF141323; NID:g5532445; I:  
A;Experimental source: strain M90T; serotype 5a  
C;Genetics:  
A;Gene: iutA

Query Match 3.7%; Score 104; DB 2; Length 732;  
Best Local Similarity 19.2%; Pred. No. 8.4;  
Matches 83; Conservative 54; Mismatches 168; Indels 128; Gaps 15;

Qy 80 LSPAYLREGGKTDFLFDPNKEPTSEERSYQSQDNN--DICGSESVADVLKQLQW-- 135

```
Db 195 LSVAYOKFGGW-----FDGNGDATLLDNTQTGLQHSNRILDMGTGTLNIDSRQLQIIT 248
QY 136 -----EWPFOELLRLREQYQ-----RE 152
Db 249 QYYSQGGDDNYGLNLGKGFSAISGSSPPYVSKGLNSDRIPTERHLLISLQYSDSDFLGQE 308
QY 153 FKNSTYSRSSVDMLYSPAKCSRDLDFGLNALLRTPDLRWNSSNAQLLLNYSCKGYNIS 212
Db 309 LVQQVYVRDESLRYPPFTVNANKQATAFSSSQDQTD---QYGMKLTLLNSQLMDGWQIIT 364
QY 213 WELGNEPNSFWKKAQISIDQLGEDFVELHKLLOKSAFONAKLYGPDIGOPRGKTVKLL 272
Db 365 WGLDAHERFTS-----NOMFFDLAQASAGGLNNHKIY--TTGRYPSYDITNL 411
QY 273 RSFLKAGGEVIDSLT---WHHYLNGRVATKEDF-----LSSDVL----- 309
Db 412 AAFLOSSYDINDIFTVSGGVRYQYTENRV---DDFIDYTOQOKIAAGKAISADAIPGGSV 468
QY 310 --DTFILSVOKILKVTKEM-----TPGKKVWILGETSSAYGGAPLLSNTFAA 354
Db 469 DYDNFLFNAGLLMHITEHQALFNFSQGVALLPDGKYGRGIYGAAVNGHPLTKSV--- 525
QY 355 GFNWLDKLGLSAQLGIEVVMRQVFFGAGNYHLVDENEPEPLDYWLSLLFKLVGPVULMS 414
Db 526 -----NVSDSKLEGVKVDSYEL-----GWRFIGDNLRTQIAAYYSLSNKSVERNKDLTI 574
QY 415 RVKGPDRSKLRVY 427
Db 575 SVKDDRR---RIY 584
```

Search completed: June 5, 2006, 12:21:46  
Job time : 20.8454 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:21 ; Search time 132.991 Seconds  
(without alignments)  
3728.138 Million cell updates/sec

Title: US-10-645-659A-3  
Perfect score: 2800  
Sequence: 1 MLAPLLLLLWGLRALTGQ.....VPAFSYGFVIRNAKIAICI 536

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt 7.2.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2783	99.4	536	1 HPSE_RAT	Q7lrlp1 rattus norv
2	2576.5	92.0	535	1 HPSE_MOUSE	Q6ygz1 mus musculu
3	2336	83.4	574	2 Q333x8	Q333x8 spalax gola
4	2336	83.4	574	2 Q333x9	Q333x9 spalax gali
5	2331	83.2	574	2 Q333x7	Q333x7 spalax carm
6	2323	83.0	574	2 Q333x6	Q333x6 spalax juda
7	2218	79.2	558	2 Q333x5	Q333x5 spalax juda
8	2151	76.8	545	1 HPSE_BOVIN	Q9myy0 bos tauru
9	2126	75.9	543	1 HPSE_HUMAN	Q9y251 homo sapien
10	1577	56.3	523	1 HPSE_CHICK	Q30yk5 gallus gall
11	1290	46.1	533	2 Q4SYF6	Q4syf6 tetradon n
12	1171	41.8	592	1 HPSE2_HUMAN	Q9wwq2 homo sapien
13	1171	41.8	592	2 Q2MIH9	Q2mih9 homo sapien
14	1048	37.4	597	2 Q4TB80	Q4tcb80 tetradon n
15	725.5	25.9	255	2 Q4TGC8	Q4tgc8 tetradon n
16	669	23.9	515	2 Q8T108	Q8t108 bombyx mori
17	410	14.6	543	1 HPSE3_ARATH	Q9ffl0 arabidopsis
18	385	13.8	559	2 Q89F99	Q89f99 bradyrhizob
19	374	13.4	527	2 Q9LR08	Q9lrc8 scutellaria
20	368.5	13.2	526	2 Q5SNA6	Q5sna6 oryza sativ
21	368.5	13.2	541	2 Q69116	Q69116 oryza sativ
22	357.5	12.8	539	2 Q2QN56	Q2qn56 oryza sativ
23	353	12.6	536	1 HPSE3_ARATH	Q9ifzp1 arabidopsis
24	352.5	12.6	537	2 Q70YJ3	Q70yj3 hordeum vul
25	350	12.5	529	2 Q6ZJE2	Q6zje2 oryza sativ
26	336	12.0	401	2 Q30324	Q30324 arabidopsis
27	330.5	11.8	539	1 HPSE2_ARATH	Q81608 arabidopsis
28	330	11.8	516	2 Q447R5	Q447r5 solibacter
29	299	10.7	537	2 Q43S03	Q43s03 solibacter
30	272	9.7	506	2 Q37Q70	Q37q70 novosphingo
31	239.5	8.6	382	2 Q3E8P7	Q3e8p7 arabidopsis

32	170	6.1	1128	2 Q5TT65	Q5tt65 anopheles g
33	154	5.5	510	2 Q2U0T3	Q2u0t3 aspergillus
34	139.5	5.0	935	2 Q9VE79	Q9ve79 drosophila
35	137.5	4.9	1053	2 P71329	P71329 fibrobacter
36	137.5	4.9	1053	2 P77865	P77865 fibrobacter
37	132.5	4.7	670	2 Q3JTG0	Q3jtg0 burkholderi
38	131.5	4.7	463	2 Q63T97	Q63t97 burkholderi
39	128	4.6	559	2 Q7SFB0	Q7sfb0 neurospora
40	125.5	4.5	536	2 Q2UDS9	Q2uds9 aspergillus
41	125	4.5	1175	2 Q5KTI5	Q5kti5 microbulbif
42	124	4.4	739	2 Q8EWI3	Q8ewi3 mycoplasma
43	123	4.4	689	2 Q5VNX0	Q5vnx0 nocardia fa
44	122.5	4.4	628	2 Q2ZEV8	Q2zev8 caldicellul
45	120	4.3	721	2 Q3XV07	Q3xv07 magnetococc

ALIGNMENTS

RESULT 1  
HPSE\_RAT ID HPSE\_RAT STANDARD; PRT; 536 AA.  
AC Q7lrlp1; Q9QZP8;  
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
DT 05-JUL-2004, sequence version 1.  
DT 07-MAR-2006, entry version 11.  
DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) (Contains:  
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit).  
GN Name=Hpse; Synonyms=Hep;  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA].  
RC TISSUE=Placenta;  
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;  
RA Hullett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,  
RA Parish C.R.;

RT Cloning of mammalian heparanase, an important enzyme in tumor  
RT invasion and metastasis.";  
RL Nat. Med. 5:803-809(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.  
RX MEDLINE=22194309; PubMed=12077130; DOI=10.1074/jbc.M203282200;  
RA Podyma-Inoue K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;  
RT "Characterization of heparanase from a rat parathyroid cell line."  
J. Biol. Chem. 277:32459-32465(2002).  
CC -|- FUNCTION: Endoglycosidase which is a cell surface and  
extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
proteoglycans (HSPGs) into heparan sulfate side chains and core  
proteoglycans. Also implicated in the extravasation of leukocytes  
and tumor cell lines. Contributes to metastasis and angiogenesis  
(By similarity).  
CC -|- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower  
extent, by heparin and sulfamin (By similarity). Activated by  
calcium and magnesium. Inhibited by EDTA.  
CC -|- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
subunits, the proteolytic products (By similarity).  
CC -|- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.  
Secreted, internalised and transferred to late endosomes/lysosomes  
as a proheparanase. In lysosomes, it is processed into the active  
form, the heparanase. The uptake or internalisation of  
proheparanase is mediated by HSPGs. Heparin appears to be a  
competitor and retain proheparanase in the extracellular medium  
(By similarity).  
CC -|- PTM: Proteolytically processed. The cleavage of the 65 kDa form  
leads to the generation of a linker peptide, 8 kDa and 50 kDa  
product. The active form, the 8/50 kDa heterodimer, is resistant  
to degradation. Complete removal of the linker peptide appears to  
be a prerequisite to the complete activation of the enzyme (By  
similarity).



CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears  
CC to be essential for its solubility (By similarity).  
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AF359508; AAQ15189.1; -; mRNA.  
DR EMBL; AF184967; AAF04563.1; -; mRNA.  
DR RGD; 61969; Hpsa.  
DR InterPro; IPR005199; Glyco\_hydro\_79\_N.  
DR Pfam; PF03662; Glyco\_hydro\_79n; I\_ N.  
KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;  
KW Signal.  
FT SIGNAL 1 28 By similarity.  
FT CHAIN 29 102 Heparanase 8 kDa subunit.  
FT PROPEP 103 150 Linker peptide (By similarity).  
FT CHAIN 151 536 /FTId:PRO\_0000042267.  
FT REGION 151 155 Heparanase 50 kDa subunit.  
FT REGION 263 273 /FTId:PRO\_0000042268.  
FT ACT\_SITE 218 218 Heparin/HS-binding (By similarity).  
FT ACT\_SITE 336 336 Heparin/HS-binding (By similarity).  
FT CARBOHYD 155 155 Proton donor (Potential).  
FT CARBOHYD 193 193 N-linked (GlcNAc...) (By similarity).  
FT CARBOHYD 210 210 N-linked (GlcNAc...) (By similarity).  
FT CARBOHYD 452 452 N-linked (GlcNAc...) (By similarity).  
FT CONFLICT 15 15 G -> R (in Ref. 2).  
FT CONFLICT 227 227 H -> Q (in Ref. 2).  
FT CONFLICT 350 350 D -> N (in Ref. 2).  
SQ SEQUENCE 536 AA; 60480 MW; C434E04CF536EA4D CRC64;  
  
Query Match 99.4%; Score 2783; DB 1; Length 536;  
Best Local Similarity 99.4%; Pred. No. 9.8e-200;  
Matches 533; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 MLRPLLLMLWGLRALTQCTPGAGTAPTKDVLDFYTKRFLFQSVSPFLSITIDASLAT 60  
DB 1 MLRPLLLMLWGLRALTQCTPGAGTAPTKDVLDFYTKRFLFQSVSPFLSITIDASLAT 60  
  
QY 61 DRPFLTFLGSPRLRALARGLSPAYLFGGKTDFLIPDPNKEPTSEERSYQSDNNNDIC 120  
DB 61 DRPFLTFLGSPRLRALARGLSPAYLFGGKTDFLIPDPNKEPTSEERSYQSDNNNDIC 120  
  
QY 121 GSERVSADVLRLKQWMPFOELLREYQREPKNSTYRSRSDMLYSFAKCSRLDLIFG 180  
DB 121 GSERVSADVLRLKQWMPFOELLREYQREPKNSTYRSRSDMLYSFAKCSRLDLIFG 180  
  
QY 181 LNALRTPLRLWSSNAQLLLNVCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240  
DB 181 LNALRTPLRLWSSNAQLLLNVCSSKGYNISWELGNEPNSFWKKAHISIDGLQGEDFV 240  
  
QY 241 ELHKLQKSAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGEVDLSLTWHYLLNGRVATK 300  
DB 241 ELHKLQKSAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGEVDLSLTWHYLLNGRVATK 300  
  
QY 301 EDPLSSDVLDTFLSLVQKILKVTKEPMPGKVMWLGSETSSAYGGAPLLSNFTFAAGFMWLD 360  
DB 301 EDPLSSDVLDTFLSLVQKILKVTKEPMPGKVMWLGSETSSAYGGAPLLSDTFAAGFMWLD 360  
  
QY 361 KLGLSQAQLGIEVVMRQVFFAGNYHLVDENFELPDYWLSSLKPLGPKVLSRVKGP 420  
DB 361 KLGLSQAQLGIEVVMRQVFFAGNYHLVDENFELPDYWLSSLKPLGPKVLSRVKGP 420  
  
QY 421 RSKLRVYLCTNYHPRYREGDITLVYLNLHNTYTKHLKLPMPFSRPVDKYLKPFSGSDG 480  
DB 421 RSKLRVYLCTNYHPRYREGDITLVYLNLHNTYTKHLKLPMPFSRPVDKYLKPFSGSDG 480  
  
QY 481 LLKSKVQLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPAFSYGFFVIRNAKIAACI 536  
DB 481 LLKSKVQLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPAFSYGFFVIRNAKIAACI 536

RESULT 2  
ID HPSE\_MOUSE STANDARD; PRT; 535 AA.  
AC Q6YGL; Q8K3K3;  
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
DT 07-MAR-2006, entry version 13.  
DE Heparanase precursor (BC 3.2.-.-) (Endo-glucuronidase) [Contains:  
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].  
GN Name=Hpsa; Synonyms=Hpa;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA].  
RC STRAIN=SJL/J; TISSUE=Spleen;  
RX MEDLINE=93321249; PubMed=10395326; DOI=10.1038/10525;  
RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,  
RA Parish C.R.;  
RT "Cloning of mammalian heparanase, an important enzyme in tumor  
RT invasion and metastasis";  
RT Nat. Med. 5:803-809(1999).  
[2]  
RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 28-57 AND 150-179,  
RP GLYCOSYLATION, BIOPHYSICOCHEMICAL PROPERTIES, ENZYME REGULATION, AND  
RP SUBUNITS.  
RC STRAIN=FVB; TISSUE=Embryo;  
RX MEDLINE=22350326; PubMed=12460766; DOI=10.1016/S1046-5928(02)00558-2;  
RA Mao H.-Q., Navarro E., Patel S., Sargent D., Koo H., Wan H.,  
RA Plata A., Zhou Q., Ludwig D., Bohlen P., Kussie P.;  
RT "Cloning, expression, and purification of mouse heparanase";  
RT Protein Expr. Purif. 26:425-431(2002).  
[3]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.  
RX MEDLINE=22841152; PubMed=12837765; DOI=10.1074/jbc.M300925200;  
RA Gong F., Jemth P., Galvis M.L.E., Vlodavsky I., Horner A., Lindahl U.,  
RA Li J.-P.;  
RT "Processing of macromolecular heparin by heparanase";  
RT J. Biol. Chem. 278:35152-35158(2003).  
[4]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC STRAIN=C57BL/6J, and NOD; TISSUE=Thymus;  
RX PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgi-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hilt M., Huminicki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Ljun S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Motagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,  
RA Nakazoni R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavoni G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Samadpour A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugiura K., Sultana R., Takenaka Y., Taki K.,

FT	CARBOHYD	230	230	N-linked (GlcNAc. . .)	(Potential).				
FT	CARBOHYD	451	451	N-linked (GlcNAc. . .) <td>(Potential).</td>	(Potential).				
FT	CONFLICT	206	206	K -> R (in Ref. 3).					
FT	CONFLICT	212	212	W -> S (in Ref. 3).					
FT	CONFLICT	230	232	NGS -> DGL (in Ref. 1, 2 and 4).					
FT	CONFLICT	335	335	E -> K (in Ref. 3).					
FT	CONFLICT	342	342	G -> A (in Ref. 3).					
FT	CONFLICT	455	455	Y -> H (in Ref. 1, 2 and 4).					
FT	CONFLICT	531	531	V -> I (in Ref. 1, 2 and 4).					
FT	SEQUENCE	535 AA;	60050 MW;	AF19B28B7CD03F7B CRC64;					
Query Match									
Best Local Similarity		92.0%; Score 2576.5;		DB 1; Length 535;					
Matches 493; Conservative		22; Mismatches		20; Indels					
		1; Gaps		1;					
QY	1	MLRPLLLWGLRRLAL	TQGT	PACTAPT	KOVVDLEFFYTKRPLFSQSPSPSLITIDASLAT				
DB	1	MLR-LLLLLWGLALQA	QAPACT	APT	DDVVDLEFFYTKRPLRSVSPSPSLITIDASLAT				
QY	61	DRPFLFLGSPRLAR	ALG	LSPAYLR	EGGTTKDFLI				
DB	60	DRPFLFLGSPRLAR	ALG	LSPAYLR	EGGTTKDFLI				
QY	121	GSRVSNVLRKLOME	WPFQELL	LLLR	EQYQREFKNSYTSRSSVDMLYSPAKCSRDLDFEG				
DB	120	RSEFVSAVLRKLQ	VEWPFQELL	LLLR	EQYQREFKNSYTSRSSVDMLYSPAKCSRDLDFEG				
QY	181	LNALLRTPDLRWNS	NAQLLN	YSCSSKGYNISWEL	GNENPNSFWKKAQISIDGLQGEDFV				
DB	180	LNALLRTPDLRWNS	NAQLLL	YSCSSKGYNISWEL	GNENPNSFWKKAHLLINGSQGEDFV				
QY	241	ELHKLQKSAFONAK	LYGPDIG	QPRGKTVKLL	RSFLKAGGEVDSLTWHYYLNGRVATK				
DB	240	ELHKLQKSAFONAK	LYGPDIG	QPRGKTVKLL	RSFLKAGGEVDSLTWHYYLNGRIATK				
QY	301	EDFLSSDVLDTFIL	SVOKIL	IKVTKE	TPGKVKWLGSETSSAYGGAPLLSNTFAAGFMWLD				
DB	300	EDFLSSDVLDTFIL	SVOKIL	IKVTKE	TPGKVKWLGSETSSAYGGAPLLSNTFAAGFMWLD				
QY	361	KLGLSAQIG	EVVMRQVFF	GAGNYHLVDEN	PEPLPDYWL				
DB	360	KLGLSAQIG	EVVMRQVFF	GAGNYHLVDEN	PEPLPDYWL				
QY	421	RSKLRYVHLCTN	VYHP	RYREGD	LTLYVLNLHNVTKHLKLPDPMFSPRPVDYKLLKPFQSGD				
DB	420	RSKLRYVHLCTN	VYHP	RYREGD	LTLYVLNLHNVTKHLKLPDPMFSPRPVDYKLLKPFQSGD				
QY	481	LLSKSVQLNGOTL	KWDEQTL	LPALTEKPL	PAGSSLSVPAFSVGFVIRNAKTAACI				
DB	480	LLSKSVQLNGOIL	KWDEQTL	LPALTEKPL	PAGSALSLSVPAFSVGFVIRNAKVAACI				
RESULT 3									
Q33X8_9NODE PRELIMINARY; PRT; 574 AA.									
ID	Q33X8_9NODE	AC							
DT	06-DEC-2005,	integrated into UniProtKB/TrEMBL.							
DT	06-DEC-2005,	sequence version 1.							
DT	07-FEB-2006,	entry version 3.							
DE	Heparanase.								
GN	Name:hpa;								
OS	Spalax golani.								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;								
OC	Muroidea; Spalacidae; Spalacinae; Spalax.								
OX	NCBI_taxid=191382;								
RN	{1}								
RP	NUCLEOTIDE SEQUENCE.								
RC	TISSUE=Kidney;								
RA	Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;								
RT	"Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene								
RL	cloning and identification of a novel splice variant.";								
RL	Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).								

[2]
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Kidney;
RC Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RA "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
RT cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085491; CAJ30018.1; -; mRNA.
SQ SEQUENCE 574 AA; 64555 MW; 48EBBFE7D0BCB34 CRC64;
Query Match 83.4%; Score 2336; DB 2; Length 574;
Best Local Similarity 83.0%; Pred. No. 3.5e+166;
Matches 445; Conservative 40; Mismatches 49; Indels 2; Gaps 1;
QY 1 MLRPLLLMLGRLRALTOGTPAGTAPTKDVVDLEFVTKRLFQSVSPSFLSITIDASLAT 60
DB 1                             :     :     :     :     :     :     :
QY 41 MLRLSLULMLGPGSLPVQCILA--AQADEVELEFSTQRPLHLVSPFSLSITIDANLAT 98
DB 1                             :     :     :     :     :     :     :
QY 61 DPRFLTGLGPRLRALARGLSPAYLRFGGTKDTFLIFDPNKPEPTSEERSYQSQDNNDIC 120
DB 1                             :     :     :     :     :     :     :
QY 99 DPRFLTGLGPKRLARALARGLSPAYLRFGGTKDTFLIFDPKKEPSHEERSYKSWNHDI 158
DB 1                             :     :     :     :     :     :     :
QY 121 GSERVSADVLRLKLQWEPFQBELLRLREQYQREFKNYSRSSSDMLYSFAKCSRDLDFG 180
DB 1 :     :     :     :     :     :     :     :     :     :     :     :
QY 159 RSGAIPAVVRRLQVENPFQQLLRREQYQKEFKNSTYSSRSVDMLYTFARCSGLDLIFG 218
DB 1     :     :     :     :     :     :     :     :     :     :     :
QY 181 LNALRRTPDLRWNSNAQLLNYSCKSGNYTSWELGNPNFSFWKAQISIDGLQGSDFFV 240
DB 1     :     :     :     :     :     :     :     :     :     :     :
QY 219 LNALRRTADFRWNSSNAQLLNYSCKSNKYDISWELGNPNFSFWKAHISIDGLQGEDYI 278
DB 1     :     :     :     :     :     :     :     :     :     :     :
QY 241 ELHKLLQSAFONAKLYGPDIGQPRGTVKLLRSFLKAGGVIDSLTWHHYLYNGRVATK 300
DB 1     :     :     :     :     :     :     :     :     :     :     :
QY 279 ELHKLLRKSTLKNNKLYGPDVGQPRGTVKLLRSFLKAGGVIDSVTWHHYLYNGRIATK 338
DB 1     :     :     :     :     :     :     :     :     :     :     :
QY 301 EDFLSSDVLDTFILSVOKILKVTKEMTPGKKVWLGETSSAYGGGAPLLSNTFAAGFWLMD 360
DB 1                             :     :     :     :     :     :     :
QY 339 EDFLSPDVLDTFILSVOKILQVSETRFGKKVWLGETSSAYGGGAPLLSNTFAAGFWLMD 398
DB 1                             :     :     :     :     :     :     :
QY 361 KGLSALOGLTEVVMNRQVFFGAGNVHLVDENPELPDYWLSSLFKLVGPKVLMSRVKGD 420
DB 1     :     :     :     :     :     :     :     :     :     :     :
QY 399 KLGLUSSQNGIEVVMNRQVFFGAGNVHLVDKNFEPLPDYWLSSLFKLVGSKVLMAVKGPD 458
DB 1     :     :     :     :     :     :     :     :     :     :     :
QY 421 RSKLRVYLHCTNVYHPRYREGDGLTYLVNLHNVTKHKLPPPMFSRPDYKLLKFPFGSDG 480
DB 1     :     :     :     :     :     :     :     :     :     :     :
QY 459 RSKLRVYLHCTINHPRYQEGDGLTYALNLYNVTKHLKLPQLFNKPDKYLVKPFLPGG 518
DB 1     :     :     :     :     :     :     :     :     :     :     :
QY 481 LLKSQVOLNGQTQKWDEQTLPALTEKPLPAGSSLSVPAFSYGFVIRNAKIAACI 536
DB 1     :     :     :     :     :     :     :     :     :     :     :
QY 519 LLSKSVOLNGOALKWDDOTPALTEKPLRPGRSSLGLPAFSYGFVIRNAKVAACL 574
DB 1     :     :     :     :     :     :     :     :     :     :     :

RESULT 4	
Q333X9_5RODE	
ID Q333X9_9RODE	PRELIMINARY; PRT; 574 AA.
AC Q333X9;	
DT 06-DEC-2005,	integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005,	sequence version 1.
DT 07-FEB-2006,	entry version 3.
DE Heparanase.	
DE Heparanase.	
GN Name=hpa;	
OS Spalax galili.	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;	
OC Muridea; Spalacidae; Spalacinae; Spalax.	
OX NCBI_TaxID=164323;	
OX [1]	
RN NUCLEOTIDE SEQUENCE.	
RP TSSUR=Kidney;	
RC RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;	

	"Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene cloning and identification of a novel splice variant.", Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0). (2)
RT	NUCLEOTIDE SEQUENCE.
RN	TISSUE=Kidney;
RC	Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RA	"Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene cloning and identification of a novel splice variant.", Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
RT	-----
RL	-----
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms Distributed under the Creative Commons Attribution-NoDerivs License -----
CC	EMBL; AM085490; CAJ30017.1; -- mRNA.
DR	SEQUENCE 574 AA; 64525 MW; 1635865051B380D0 CRC64;
SQ	
	Query Match 83.4%; Score 2336; DB 2; Length 574; Best Local Similarity 83.0%; Pred. No. 3.5e-166; Matches 445; Conservative 40; Mismatches 49; Indels 2; Gaps 1
Qy	1 MLRPLLLLLLWGRALRALTQGTPAGTAPTKDVVVDLFEYFKRLFQSVPSPFLSITIDASLAT 60
Db	41 MLRUSLLLLLWGPISLPVQCILA--AQADEVVELEFSTQRPLHLVSPSPFLSITIDANLAT 98
Qy	61 DPRFLTFLGSPRLRALARGLSPAYLRFGGTKTDFLIFDPNKPEPTSEERSYMOSQNNDIC 120
Db	99 DPRFLTFLGSPKLRALARGLSPAYLRFGGTKTDFLIFDPKKEPSHEERSYWKSVQNHDC 158
Qy	121 GSERVSADVLRKLQWEHPFPQELLLLREOYQREFKNSTYSRSSVDMLXSFPAKCSRLLDLIFG 180
Db	159 RSGAIPAVVRRILQVWPFPQQLIRLEOYQDFKNSTYSRSSVDMLYTFARCSGLDLIFG 218
Qy	181 LNALLRTPDLRWNSSNQALLLNYSCKGYNTSWELGNPNPSFWKKAQISIDLGLQGEDVF 240
Db	219 LNALLRADFWNSNAQALLLNYSCKKYDIWSWELGNPNPSFWKKAHISIDLGLQGEYI 278
Qy	241 ELHKLLQKSAPONAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLYNGRVATPK 300
Db	279 ELHKLLRKSTLKNNVKLYGPDVGQPRGKTVKLLRSFLKAGGEVIDSVTWHYYLYNGRIATP 338
Qy	301 EDFLSSVDLDTFILSVOKILKVTEMTPGKKVMJGETSSAYGGGAPLLSNTFAAGFMWLD 360
Db	339 EDFLSPDVDLDTFILSVQKILQVETREPGKKVMJGETSSAYGGGAPLLSNTFAAGFMWLD 398
Qy	361 KLGLSAQLEGIEVNMQRVPFGAGNYHLVDENPEPLPDYWLSSLFKLVGPKVLMSRVKGPD 420
Db	399 KLGLSAQMGIEVNMQRVPFGAGNYHLVDKNFEPLPDYWLSSLFKLVGSKVLMAVRVKGPD 458
Qy	421 RSKLRVYLHCNVVHPRVREGDLTLVYLNHNVTKHKLPPDMFSRPVDKYLKLPFGSDG 480
Db	459 RSKURVYLHCNTINHPRYQEGBLTLYALNLNVYTKHLKPQLFNKKPNVDKVLVPLGPGG 518
Qy	481 LLSSKSVQLNGOTLKMVBDEQTLPALTEKPLPAGSSLSVPAFSYGFVFINAKTAACI 536
b	519 LISSKSVOLNGOAIKMVDODTLPALTEKPLRGSSLGLPAFSYGFVFINAKVAACL 574

RESULT 5	
Q333X7_9GRADE	
Q333X7_9CODE	PRELIMINARY; PRT; 574 AA.
ID	
AC	Q333X7_
DT	06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT	06-DEC-2005, sequence version 1.
DT	07-FEB-2006, entry version 3.
DE	Heparanase.
GN	Name=hpa;
OS	Spalax catmelli.
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC	Muroidea; Spalacidae; Spalacinae; Spalax.
OX	NCBI_TaxID=164324;
RN	[1]

```
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085492; CAJ30019.1; -, mRNA.
SQ SEQUENCE 574 AA; 64459 MW; 9F1D19DCB99DE CRC64;

Query Match      83.28; Score 2331; DB 2; Length 574;
Best Local Similarity 83.0%; Pred. No. 8.4e-166;
Matches 445; Conservative 39; Mismatches 50; Indels 2; Gaps 1;

Qy 1 MLRPLLLLMWGRRLALTOGTPTAGTAPTKDQVVDLEFYTKLRFQSVSPSFLSITIDASLAT 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41 MLRLSLLLMWGPLSPVQCILA--AQAEVVELEFSTQRPDLHLVSPSFLSITIDANLAT 98
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 61 DPRLFTFLGSPRLARALARGLSPAYLRFGGTKTDFLIDPNKEPTSEERSYMQSDNNDIC 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99 DPRLFTFLGSPKRLARALARGLSPAYLRFGGTKTDFLIDPKKEPSHEERSYMQSDNNDIC 158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 121 GSERVSADVLRLKQMEWPPFQELLRLRLREYQREFKNSTYSRSSVDMLYSPAKCSRLDLIFG 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 RSGAIPAVVVRRLQVWEPFQELLRLRLREYQREFKNSTYSRSSVDMLYTFARCSGLDLIFG 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 181 LNALLRTPDLRWSSNAQLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 LNALLRTADFRWSSNAQLLNYCSSKGYNISWELGNEPNSFWKKAHISIDGLQGEDYI 278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 241 ELHKLKQSAFONAKLYGPDIGPRGKTVKLLRSFLKAGGEVIDSLTWHHYLYNGRVATK 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 279 ELRKLARKSTLKNVLYGPDVGPRGKTVKLLRSFLKAGGEVIDSVTWHHYLYNGRIATK 338
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 301 EDFLSSDVLDTFILSVOKILKVTKEMTGKVKWLGTSAYGGAPLLSNTFAAGFWMLD 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 EDFLSPDLDTFILSVOKILQVVEETRPKKVWLGTSAYGGAPLLSNTFAAGFWMLD 398
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 361 KLGLSAQLGIEVVMRQVFFGAGNYHLVDENFEPLPDYWLSSLFKKLVGPKVMSRVKGPD 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 399 KLGLSAQMGIIEVVMRQVFFGAGNYHLVDKNFEPLPDYWLSSLFKKLVGSKVLMARVKGPD 458
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 421 RSKLRVYLHCTNVTHPRYREGDLTYVNLNHNVTKHLLKPPMFSRPVDKYLKPFPGSDG 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 459 RSKLRVYLHCTNINHPRYQEGDLTYALNLYNVTKHLLPYQLFNKPVDPKYLVIPLPGGG 518
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 481 LLSKSVQLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPFASVYGFVIRNAKIAACI 536
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 519 LLSKSVQLNGQALKMVDQTLPALTEKPLRPGSSLSGLPFAFSYGFVIRNAKVAACL 574
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
Q333X6 SPAJD
ID Q333X6 SPAJD PRELIMINARY; PRT; 574 AA.
AC Q333X6
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name:hpa;
OS Spalax judaei (Blind subterranean mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=134510;
RN [1]
RC NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
```

```
cloning and identification of a novel splice variant.";
Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
[2]
RC NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AM085493; CAJ30020.1; -, mRNA.
SQ SEQUENCE 574 AA; 64515 MW; 3AEBB13F07451684 CRC64;

Query Match      83.0%; Score 2323; DB 2; Length 574;
Best Local Similarity 82.8%; Pred. No. 3.3e-165;
Matches 444; Conservative 39; Mismatches 51; Indels 2; Gaps 1;

Qy 1 MLRPLLLLMWGRRLALTOGTPTAGTAPTKDQVVDLEFYTKLRFQSVSPSFLSITIDASLAT 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41 MLRLSLLLMWGPLSPVQCILA--AQAEVVELEFSTQRPDLHLVSPSFLSITIDANLAT 98
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 61 DPRLFTFLGSPRLARALARGLSPAYLRFGGTKTDFLIDPNKEPTSEERSYMQSDNNDIC 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99 DPRLFTFLGSPKRLARALARGLSPAYLRFGGTKTDFLIDPKKEPSHEERSYMQSDNNDIC 158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 121 GSERVSADVLRLKQMEWPPFQELLRLRLREYQREFKNSTYSRSSVDMLYSPAKCSRLDLIFG 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 RSGAIPAVVVRRLQVWEPFQELLRLRLREYQREFKNSTYSRSSVDMLYTFARCSGLDLIFG 218
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 181 LNALLRTPDLRWSSNAQLLNYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 LNALLRTADFRWSSNAQLLNYCSSKGYNISWELGNEPNSFWKKAHISIDGLQGEDYI 278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 241 ELHKLKQSAFONAKLYGPDIGPRGKTVKLLRSFLKAGGEVIDSLTWHHYLYNGRVATK 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 279 ELRKLARKSTLKNVLYGPDVGPRGKTVKLLRSFLKAGGEVIDSVTWHHYLYNGRIATK 338
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 301 EDFLSSDVLDTFILSVOKILKVTKEMTGKVKWLGTSAYGGAPLLSNTFAAGFWMLD 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 EDFLSPDLDTFILSVOKILQVVEETRPKKVWLGTSAYGGAPLLSNTFAAGFWMLD 398
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 361 KLGLSAQLGIEVVMRQVFFGAGNYHLVDENFEPLPDYWLSSLFKKLVGPKVMSRVKGPD 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 399 KLGLSAQMGIIEVVMRQVFFGAGNYHLVDKNFEPLPDYWLSSLFKKLVGSKVLMARVKGPD 458
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 421 RSKLRVYLHCTNVTHPRYREGDLTYVNLNHNVTKHLLKPPMFSRPVDKYLKPFPGSDG 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 459 RSKLRVYLHCTNINHPRYQEGDLTYALNLYNVTKHLLPYQLFNKPVDPKYLVIPLPGGG 518
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 481 LLSKSVQLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPFASVYGFVIRNAKIAACI 536
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 519 LLSKSVQLNGQALKMVDQTLPALTEKPLRPGSSLSGLPFAFSYGFVIRNAKVAACL 574
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
Q333X5 SPAJD
ID Q333X5 SPAJD PRELIMINARY; PRT; 558 AA.
AC Q333X5
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name:hpa;
OS Spalax judaei (Blind subterranean mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=134510;
RN [1]
RC NUCLEOTIDE SEQUENCE.
```



Db 69 TDRPFTFLGSSKLTTLARGLAPAYLRFNGKNGDFLIFDPKKEPAFEERSYMLSQSQNDI 128  
QY 120 CGSERVSADVLRLKLOMPPQOELLRLREOYQREBFKNSTYSRSSYDMLYSPAKCSRDLIF 179  
Db 129 CKSGSPSDVEEKLRLLEWPFQEQVLLREOYQKKFTNSTYSRSSYDMLYTFASCGLNLIJ 188  
QY 180 GLNALLRTDPLRWNSSNAQLLNLCSSGKYNISWELGNEPNSFWKKAQISIDGLQLGDF 239  
Db 189 GVNALLRTDTHWDSSNAQLLNLCSSKNTNISWELGNEPNSFORAKGIFINGQLGDF 248  
QY 240 VELHLKLQKSAFONAKLYGPDIGOPRGKTIVKLRSFLKAGGEVIDSLTWHYYLNGRVAT 299  
Db 249 IEFKLLGSAFNKAKLYGPDIGOPRENTVTKMLKSLFKAGGEVIDSVTWHYYNGRIAT 308  
QY 300 KEDFLSDVLDITFLSVQKILKVTYKMPGKVKWLGTSSTAYGGAPLLSNTPAAGFWL 359  
Db 309 KEDFLNPDIIDTFTISSVQKTLRIVEKIRPLKVKWLGTSSTAYGGAPFLSNTPAAGFWL 368  
QY 360 DKLGSLAQGLIEVVMROVFFGAGNYHLVDENFELPDYWLSSLFELKVLGPKVLMRSVKGP 419  
Db 369 DKLGSLARMGIEVVMRQVLFAGNYHLVDGNFELPDYWLSSLFELKVLGPKVLMRSVKGP 428  
QY 420 DRSKRLVYLCTNVYHPRYREGDITLVLMHNVTYKHLPPPMFSRPVDKYLKLPFGSD 479  
Db 429 DRSKFRVYLCTNWKHPRYREGDITLVLMHNVTYKHLPLPHLNFQVQDKYLKLPFGSD 488  
QY 480 GLLSKSVQLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPFYSFFVIRNAKIAACI 536  
Db 489 GLLSKSVQLNGQTLKMWDEQTLPALTEKPLPAGSSLSGMPFYSFFVIRNAKVAACI 545

## RESULT 9

HPSE HUMAN  
ID HPSE HUMAN STANDARD; PRT; 543 AA.  
AC Q9Y2E1: Q53GE5: Q9UL39;  
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
DT 01-NOV-1999, sequence version 1.  
DE Heparanase precursor (EC 3.2.-.-) (Heparanase-1) (Hpal) (Endo-glucuronidase) [Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].  
GN Names=HPSE; Synonyms=HBP, HPA, HPA1, HPRI, HPSE1, HSE1;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.  
RC TISSUE=Placenta;  
RX MEDLINE=99335379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;  
RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C., Seddon A.P., Giorgio N.A., Bohlen P.;  
RT "Cloning and functional expression of a human heparanase gene.";  
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN SEQUENCE OF 158-168; 326-337 AND 447-491.  
RC TISSUE=Embryonic fibroblast;  
RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;  
RA Toyoshima M., Nakajima M.;  
RT "Human heparanase. Purification, characterization, cloning, and expression.";  
RL J. Biol. Chem. 274:24153-24160(1999).  
RN [3]  
RP NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.  
RX PubMed=10395325; DOI=10.1038/10518;  
RA Vlodavsky I., Friedmann Y., Elkin M., Aingorn H., Atzmon R., Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I., Spector L., Pecker I.;  
RT "Mammalian heparanase: gene cloning, expression and function in tumor progression and metastasis.";

Nat. Med. 5:793-802(1999).  
RN [4]  
RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE OF 158-174; 263-272; 326-337; 433-436; 438-443; 466-468 AND 478-483.  
RC TISSUE=Placenta;  
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;  
RA Hultett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J., Parish C.R.;  
RT "Cloning of mammalian heparanase, an important enzyme in tumor invasion and metastasis.";  
RN Nat. Med. 5:803-809(1999).  
RN [5]  
RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE SPECIFICITY.  
RC TISSUE=Placenta;  
RX MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;  
RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;  
RT "Heparanase expression in invasive trophoblasts and acute vascular damage.";  
RL Glycobiology 10:467-475(2000).  
RN [6]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.  
RX PubMed=11547900; DOI=10.1023/A:1011375624902;  
RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O., Aviv A., Elkin M., Pecker I., Peretz T., Vlodavsky I.;  
RT "Molecular properties and involvement of heparanase in cancer progression and mammary gland morphogenesis.";  
RN J. Mammary Gland Biol. Neoplasia 6:311-322(2001).  
RN [7]  
RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163, SUBUNTS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.  
RC TISSUE=Placenta;  
RX PubMed=12713442; DOI=10.1042/BJ20030318;  
RA McKenzie E., Young K., Hircok M., Bennett J., Bhaman M., Felix R., Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S., Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M., Parekh R., Stubberfield C.;  
RT "Biochemical characterization of the active heterodimer form of human heparanase (Hpal) protein expressed in insect cells.";  
RL Biochem. J. 373:423-435(2003).  
RN [8]  
RP NUCLEOTIDE SEQUENCE [MRNA].  
RA Pinhal M.A., Semedo P.;  
RT "Cloned heparanase from MCF-7 cells.";  
RN Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.  
RN [9]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC TISSUE=Small intestine;  
RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y., Tanaka A., Yokoyama S.;  
RN Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.  
RN [10]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC TISSUE=Pancreas;  
RX MEDLINE=25388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udgin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human







```
RX MEDLINE=21369959; PubMed=11387326; DOI=10.1074/jbc.M102462200;
RA Goldehmidt O., Zcharia E., Aingorn H., Guatta-Rangini Z., Azmon R.,
RA Michal I., Pecker I., Mitran E., Vlodavsky I.;
RT "Expression pattern and secretion of human and chicken heparanase are
RT determined by their signal peptide sequence.";
RL J. Biol. Chem. 276:29178-29187(2001).
CC -I- FUNCTION: Endoglycosidase which is a cell surface and
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC proteoglycans (HSPGs) into heparan sulfate side chains and core
CC proteoglycans (by similarity).
CC -I- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted
CC (by similarity).
CC -I- PTM: N-glycosylated (by similarity).
CC -I- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: AY037007; AAK82648.1; -; mRNA.
DR Ensembl: ENSGALG0000011203; Gallus gallus.
DR InterPro: IPR005199; Glyco_hydro_79_N.
DR Pfam: PF03662; Glyco_hydro_79n; 1.
KW Glycoprotein; Hydrolase; Lysosome; Membrane; Signal.
FT SIGNAL 1 18 Potential.
FT CHAIN 19 523 Heparanase.
FT FTID=RS_0000042259.
FT REGION 137 141 Heparin/HS-binding (by similarity).
FT REGION 250 260 Heparin/HS-binding (by similarity).
FT ACT_SITE 204 204 Proton donor (potential).
FT ACT_SITE 323 323 Nucleophile (potential).
FT CARBOHYD 141 141 N-linked (GlcNAc...) (potential).
FT CARBOHYD 196 196 N-linked (GlcNAc...) (potential).
FT CARBOHYD 436 436 N-linked (GlcNAc...) (potential).
FT CARBOHYD 439 439 N-linked (GlcNAc...) (potential).
SQ SEQUENCE 523 AA; 58386 MW; 8EB0B7B18C9BF881 CRC64;

Query Match 56.38; Score 1577; DB 1; Length 523;
Best Local Similarity 59.18; Pred. No. 2.5e-109;
Matches 303; Conservative 86; Mismatches 118; Indels 6; Gaps 3;

QY 27 PTKDVVDLEFYTKLFOVSFSLITDASLATDPRFLTFLGSPRLALARGLSPAYLR 86
DB 14 PPRTAELQGLREPIGAVSPAFSLTLDASLARDPRFVALLRPKLTHTLASGLSPGLR 73
QY 87 FGCTKTDFLFDNKNKTSBER--SYWQSQDNNDICGSERSVADVLRLQWEPFQELLL 144
DB 74 FGCTSTDFLFDNKNKTSBER--SYWQSQDNNDICGSERSVADVLRLQWEPFQELLL 130
QY 145 LREQYQREFKNTYSRSDMLYSFAKCSRLDLIFGLNALLRTPDLRWNSNAQLLNYC 204
DB 131 LAESHKWKCHKNTTITRSTLDILHTFASSGFRLLVFGNALLRRAGLQWDSNAKQLLYC 190
QY 205 SSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFVELHKLK-QKSAFQNAKLYGPDIG 263
DB 191 AQRSYNISWELGNEPNSFRKXSGICDGFQGLRDFVHLRQLLSQHPYRAELYGLDVQG 250
QY 264 PRGKTVKLLRSFLKAGEVIDSLTWHYLYNGRVATKEDFLSSDVLDTFSLVQKILKVT 323
DB 251 PRKHTOHLRSFWKSGKALDSVTHWHYLYNGRSATREDFLSPVLDSFATAHDVLGIV 310
QY 324 KEMTPGKKVWLGTSAYGGAPLLSNTFAAGFMWLDKGLSQAOLGIEVVMRQVFFGAGN 383
DB 311 EATVPKVKVWLGTSAYGGAPQLSNTYVAGFMWLDKGLAARRGIDVVMRQVFFGAGS 370
QY 384 YHVLVDNFEPLPDYWLISLLPKVLGPKVMSRVKGPDRSKLRYLHCTNVTHPRYREGDL 443
DB 371 YHVLVDAGFKPLPDYWLISLLKRLVGTVLQASVEQADARRPRVYLHCTNPRHPKRYG 430
QY 444 TLVYLNLHNVTKHLKPPMFPSPVDKYLKPGSGDGLLSKSVOLNGQTLKMWDEQTLPA 503
DB 431 TLFALNLSNVTQSLQPKQLWSKSDVQYLLPHGKDSILSREVQLNGRLQMVDDTLPA 490
QY 504 LTKPLPAGSSLSVPAPSYGFFVIRNAKIACI 536
```

```
DB 491 LHMALAPGSLTGLUPAFSGYGFVIRNAKIACI 523

RESULT 11
Q4SYF6 TETNG
ID Q4SYF6_TETNG PRELIMINARY; PRT; 533 AA.
AC Q4SYF6_
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Chromosome undetermined SCAP12073, whole genome shotgun sequence.
DE (Fragment).
DE ORFNames=GSTENG00010356001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Bismont V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Craud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -I- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: CAAE01012073; CAF94326.1; -; Genomic_DNA.
FT NON_TER 1 533
FT NON_TER 533 533
SQ SEQUENCE 533 AA; 60100 MW; 9B00A7C8780100FF CRC64;

Query Match 46.11; Score 1290; DB 2; Length 533;
Best Local Similarity 48.71; Pred. No. 8.6e-88;
Matches 258; Conservative 95; Mismatches 139; Indels 38; Gaps 8;

QY 45 VSPFLSTITDASLATDPRFLTFLGSPRLALARGLSPAYLRFGCTKTDFLFDNKN-EP 103
DB 4 VDPFLSVTITDASLAADERFMYLSSPKVRTAKALTAFILRFGTQRQDMVFAPHKQF 63
QY 104 TS--EERSYWSQDNNDICGSERSVADVLRLQWEPFQELLLRLLEQYQREFKNTYSRS 161
DB 64 ASGSARELTFSSNGQSHCEKMAAPPWLERLRLETMMKQQVWLRNEELQKRYRVKFTET 123
QY 162 SVDMLYSFAKCSRLDLIFGLNALLRTPDLRWNSNAQLLNYCSSKGYNISWELGNEPNS 221
DB 124 TVDQLHAFANCSGLDLVFGNALLRTADNRWNSNARSLLRYCEARRVHMSWELGNEPNS 183
QY 222 FWKKAQISIDGLQGEDFVELHKLKQSAF-QNAKLYGPDIGQPRGKTVKLLRSFLKAGG 280
DB 184 YEKKAGRLRLDQLGEDFTVLRKILRESFRYDAGLFGDVGQPRDHRIDILSGFLQSGA 243
```



DR EMBL; AL445251; CAI17162.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAH70450.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAH70450.1; -; Genomic\_DNA.  
DR EMBL; AL139243; CAH70450.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356220; CAH70450.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAH70450.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAH70450.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAH70450.1; -; Genomic\_DNA.  
DR EMBL; AL39243; CAI16474.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356220; CAI16474.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAH70450.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAI16474.1; JOINED; Genomic\_DNA.  
DR EMBL; AL139243; CAI16474.1; -; Genomic\_DNA.  
DR EMBL; AL356220; CAI14148.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAI14148.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAI14148.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAI14148.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAH73138.1; -; Genomic\_DNA.  
DR EMBL; AL139243; CAH73138.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356220; CAH73138.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAH73138.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAH73138.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAH70449.1; -; Genomic\_DNA.  
DR EMBL; AL139243; CAH70449.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356220; CAH70449.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAH70449.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAH70449.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAI16473.1; -; Genomic\_DNA.  
DR EMBL; AL139243; CAI16473.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356220; CAI16473.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAI16473.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAI16473.1; JOINED; Genomic\_DNA.  
DR EMBL; AL139243; CAI14147.1; -; Genomic\_DNA.  
DR EMBL; AL356220; CAI14147.1; JOINED; Genomic\_DNA.  
DR EMBL; AL356268; CAI14147.1; JOINED; Genomic\_DNA.  
DR EMBL; AL445251; CAI14147.1; JOINED; Genomic\_DNA.  
DR EMBL; AL590036; CAI14147.1; JOINED; Genomic\_DNA.  
DR PIR; JC7506; JC7506.  
DR Ensembl; ENSG00000172987; Homo sapiens.  
DR HGNC; HGNC:18374; HPSE2.  
DR GO; GO:0005622; C.intracellular; TAS.  
DR GO; GO:0030305; F.inepharase activity; TAS.  
DR InterPro; IPR005199; Glyco\_hydro\_79\_N.  
DR Pfam; PF03662; Glyco\_hydro\_79n; 1.  
KW Alternative splicing; Hydrolase; Membrane; Polymorphism.  
FT CHAIN 1 592  
FT PIR; JC7506; JC7506.  
FT ACT\_SITE 262 262  
FT VARSPPLIC 150 261  
FT  
FT VARSPPLIC 204 261  
FT  
FT VARSPPLIC 539 548  
FT  
FT VARSPPLIC 549 592  
FT  
FT VARIANT 579 579  
FT  
FT CONFLICT 12 12  
FT CONFLICT 213 213  
FT SEQUENCE 592 AA; 66580 MW; 95C384AD9A74258E CRC64;  
Query Match 41.8%; Score 1171; DB 1; Length 592;  
Best Local Similarity 45.3%; Pred. No. 8.4e-79;  
Matches 249; Conservative 87; Mismatches 186; Indels 28; Gaps 7;  
QY 12 GLRLALTOGTACTACTKDVVDLEFTYKRLFSQVSPSFLSITDASLATDPRFLTLGSP 71

Db 42 GDRRLPVDRAAG-LKEKTLILLDVSTKNPVRTVNENFSLQLDPSIIHD-GWLDLFLSK 99  
QY 72 RLRLALGLSPAYLRFCTKTDFLIF-----DPNKEPTSEERSYQSQDNNDICGSE---- 123  
Db 100 RLVTTLARGLSPAFRLFGCKRTDFLQFQNLRLNPAKSRGPGPDYLLKNYEDIVRSDVALD 159  
QY 124 -----RVSADVLRLQLQWEPFO-ELLLLREQOREPKNSTYSSRSSVDMLYSPAKCSRL 175  
Db 160 KQKCKTAQHPDVMLEQREKAAQMHVLLVLEQESNTYSNLLTARSLDKLYNFADCSGL 219  
QY 176 DLIFGLNALLRTPLDRWNSSNAQLLNYCSKGYNISWELGNEPNSFWKQAQISIDGQL 235  
Db 220 HLIFALNALARNPNNSWSSALSLLKYSAKCKNISWELGNEPNNYRTMHGRAVNGSOL 279  
QY 236 GEDFVELHKLQK-SAFQNAKLYGPDIGOPRGKTVKLLRSFLKAGGEVDSLTWHYYLN 294  
Db 280 GKDYIQLKSLQPIRIYSRASLYGNIGRPRKNIALLDGFMKVAGSTVDVATWQHCV 339  
QY 295 GRVATKEDFLSSVDLDTFILSVQKILKVTXEMTPGKVKWLGTSAYGGAPLLSNTFAA 354  
Db 340 GRVVKVMDFLTKRLDLSQIRKIQKVNTYTPGKKIWLGVVTSAGTNNLSDSVA 399  
QY 355 GFWMLDKGLSAQLGIEVVMRQVFFGAGNYHLVDENFEPLPDYWLSSLFKKLVGPKVLS 414  
Db 400 GFLWNLTLGLANQCIDVIRHSFEDHGYNLVDQNFPLPDYWLSSLYKRLIGPKVLAV 459  
QY 415 RVKQPD-----RSKLRYVLHCTNVYHPRYREGDLTLVYLNHLNVTKHLKLPMPFS 465  
Db 460 HVAGLQKRPGRVIRDKLRIYAHCTNHNHNVYRGVSITLFIINLHRSKKIKLAGTLRD 519  
QY 466 RPVDKYLKPFSGDGLLSKSVQLNGQTLKMWDEQTLPALTEKPLPAGSSLSVPASYGPF 525  
Db 520 KLHGYLLQPYQGBGLSKSVQLNGQPLVMVDDGTLPELKPRPLRAGRTLVIPTVMGPF 579  
QY 526 VIRNAKIAAC 535  
Db 580 VVRNVNALAC 589  
RESULT 13  
Q2M1H9 HUMAN  
ID Q2M1H9\_HUMAN PRELIMINARY; PRT; 592 AA.  
AC Q2M1H9;  
DT 21-FEB-2006, integrated into UniProtKB/TrEMBL.  
DT 21-FEB-2006, sequence version 1.  
DT 21-FEB-2006, entry version 1.  
DE Heparanase 2.  
GN Name=HPSE2;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=PCR rescued clones;  
RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udutin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McKeown P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards D.K., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón S., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,

RA	Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human RT and mouse cDNA sequences."; Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002). [2]
RN	NUCLEOTIDE SEQUENCE.
RP	TISSUE=PCR rescued clones;
RG	NIH MGC Project;
RL	Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases. -----
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms Distributed under the Creative Commons Attribution-NoDerivs License -----
DR	EMBL; BC112356; AA112357.1; -; mRNA.
DQ	SEQUENCE 592 AA; 66610 MW; 94689E1C2A74359F CRC64;
Query Match	41.8%; Score 1171; DB 2; Length 592;
Best Local Similarity	45.3%; Pred. No. 8.4e-79;
Matches 249; Conservative 87; Mismatches 186; Indels 28; Gaps 71	
QY	12 GLRLALTTGGTPAGTAPTADVVDFYTKRLFQSVPSPFLSITIDASLATDPRLFTFLGSP 71
Db	42 GDRRPLVDRAAG-LKEKTLILDVSTKNPVRTNENFLSLQLDPSIHD-CWLDFFLSKK 99
QY	72 RLRLARGLSPAYLRFGTKTDPLIF----DPNKPTSEERSYQSQDNNDICGSE---- 123
Db	100 RLVTLAGRGLSPAFRLFPGCKRTDFLFQOWLRNPFAKSRRSGPGPDYYLKNYEDDIVRSDVALD 159
QY	124 -----RVADVLVRKLQMEWPFQ-EILLIRIQVREPKNSTYSRSSVDMLYSFAKCSRL 175
Db	160 KQGCKCTIAQHPDVMLELQREKAAQMHLVLLKEQFSNTYSNLILTFARSJDLYNFADCSSL 219
QY	176 DLIFGLNALRLTDLRWNSSNAQLLLNYCSSKGYNISWELGNPNFSFWKKAQISIDGLQL 235
Db	220 HLI FALMALARNPNNSNSSALSLLKYSAKKTNISWELGNPNPYTMGRVANGSQL 279
QY	236 GEDFVELHKLLQK-SAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGEVIDSLTWHHYYLN 294
Db	280 GKDYIQLKSLLOPIRIYSRASLYGPNIGRPKNVIALLDGFMKVAGSTVDAVTWHQCVID 339
QY	295 GRVATKEDFISSDVDLTFFILSVQIKLVKTKMTGPKVKWLGETSSAYGGGAPLISNTFAA 354
Db	340 GRVVKVMDFLKRLLDLTSDQIRKIQKVNTYTTPCKKITWLEGVVTTSAGGTNNLSDSYAA 399
QY	355 GFWMLDKLGLSAQGLIEVWRQVFEGAGNYHLDVENPEPLPDYWLISLLFKLVGPVKVIMS 414
Db	400 GFLWLNTGLMANOGIDVVIKHSPFDHGYNHLDQNFPDLWLSLKYKLLIGPKVLAV 459
QY	415 RVAGPD-----RSKLRVYLCHTNVYHPRYREGDILTLYVLNLHNVTRELKLPMPMFS 465
Db	460 HVAGLQRKPGRVIRDKLRVIAHCTNHNNHNVYRGSTLFIINLHSRKIKLTGTLRD 519
QY	466 RPVDKYLKPPGSDGLSKSVOLNGOQLKWVDEOTLPALTETKPLPAGSSLSVPFASFYGF 525
Db	520 KLHVQYLQPVQGBGLSKSVOLNGOQLVMVDDGTLPKLPKRLRAGRTRLVIPPTMGFF 579
QY	526 VIRNAKTAAC 535
Db	580 WKVNVALAC 589
RESULT 14	
Q4TB80_TETNG	PRELIMINARY; PRT; 597 AA.
ID Q4TB80_TETNG	
AC Q4TB80;	
DT 19-JUL-2005,	integrated into UniprotKB/TREMBL.
DT 19-JUL-2005,	sequence version 1.
DT 07-FEB-2006,	entry version 4.
DE Chromosome 17 SCAP180,	whole genome shotgun sequence. (Fragment).
OS ORfNames+GSTENG0003868001;	
GN Tetradodon nigroviridis (Green puffer).	
OC Eukaryota; Metazoa; Chordata; Craniata;	Vertebrata; Euteleostomi;
OC Actinopterygii; Neoterevgii; Teleostei;	Euteleostei; Neoteleostei;

[illegible]

```
Db 500 VRSIIYIINLHRSRKKIKLACTLRNNIVHQLQYPGADGLRAKHVQLNGEKLIMADN 559
Qy 499 QTPALTEKPLPAGSSLSVPFSGFFVIRNAKIAAC 535
Db 560 ETPPELKPTRLRAGRTIAMPMTIGTFYVIKININAYAC 596

RESULT 15
Q4TGC8_TETNG
ID Q4TGC8_TETNG PRELIMINARY; PRT; 255 AA.
AC Q4TGC8;
DT 19-JUL-2005, integrated into UniprotKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Chromosome undetermined SCAF3783, whole genome shotgun sequence.
DE (Fragment).
DE ORFNames=GSTENG00001168001;
GN Tetraodon nigroviridis (Green puffer).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontoidea; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poullain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Boeak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAAE01003783; CAF88054.1; -; Genomic_DNA.
FT NON_TER 1 1
FT NON_TER 255 255
SQ SEQUENCE 255 AA; 28562 MW; 07F542A9C755E3F0 CRC64;

Query Match 25.9%; Score 725.5; DB 2; Length 255;
Best Local Similarity 54.7%; Pred. No. 6.5e-46;
Matches 139; Conservative 43; Mismatches 55; Indels 17; Gaps 3;

Qy 159 SRSSVDMLYSFACSRDLIFGLNALLRTPDLRWNSSNAQLLNYCSSKGYNISWEIGNE 218
Db 1 SETTVQLHAFACNSGLDLVFLGNALLRTADNRWSSNARSLLRYCARRYHMSWEIGNE 60
Qy 219 PNSFWKKAQISIDGLQIGDFVELKHLQKSAF-ONAKLYGPDIGQPRGKTVKLLRSFLK 277
Db 61 PNSYEKKAGLRLDGRQIGEDFTVLRKILRESRFRDAGLFGPDVGQDRHRIDILSGFLQ 120
Qy 278 AGGEVIDSLTWHYHLNGRVATKEDFLSSDVLDTFILSVQILKLVTKEMTPGKKVWLGET 337
Db 121 SGAEAVDACTWHYHLDGRASLEDFLDPDVLDTLRKIGIEVLEEVHQSFGKPVWLGET 180
```

```
Qy 338 SSAYGGG-----APLLSNTFAAG-PMWLDKLGLSAQLGIEVVMRQVFFCA 381
Db 181 SSATGABPRGCRTHSSQDSCBFAPRRSQOAPLGTFRWLDKLGLAATLGLELVMRQVLGA 240
Qy 382 GNYHLVDENPEPLP 395
Db 241 GSYHLMDDNLDPLP 254
```

Search completed: June 5, 2006, 12:20:15  
Job time : 134.991 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:21 ; Search time 132.743 Seconds  
(without alignment)  
3728.138 Million cell updates/sec

Title: US-10-645-659A-2  
Perfect score: 2797  
Sequence: 1 MLRLLLMLWGLGALQA.....LPAFSYGFFVIRNAKIAICI 535

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot\_7.2.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2775	99.2	535	1 HPSE_MOUSE	Q5ygz1 mus musculus
2	2601.5	93.0	536	1 HPSE_RAT	Q71rpl rattus norv
3	2392.5	85.5	574	2 Q333x8_GRODE	Q333x8 spalax gola
4	2392.5	85.5	574	2 Q333x9_GRODE	Q333x9 spalax gali
5	2387.5	85.4	574	2 Q333x7_GRODE	Q333x7 spalax carm
6	2379.5	85.1	574	2 Q333x6_SPAJD	Q333x6 spalax juda
7	2274.5	81.3	558	2 Q333x5_SPAJD	Q333x5 spalax juda
8	2184	78.1	545	1 HPSE_BOVIN	Q9myy0 bos taurus
9	2149	76.8	543	1 HPSE_HUMAN	Q9y251 homo sapien
10	1605	57.4	523	1 HPSE_CHICK	Q90yk5 gallus gall
11	1289	46.1	533	2 Q4SF6_TETNG	Q4syf6 tetradon n
12	1169.5	41.8	592	1 HPSE2_HUMAN	Q9wq2 homo sapien
13	1169.5	41.8	592	2 Q2MIH9_HUMAN	Q2mlh9 homo sapien
14	1042.5	37.3	597	2 Q4TB80_TETNG	Q4tb80 tetradon n
15	724.5	25.9	255	2 Q4TG08_TETNG	Q4tg08 tetradon n
16	671	24.0	535	2 Q8T108_BOMMO	Q8tl08 bombyx mori
17	415	14.8	535	2 Q89F99_BRAJA	Q89f99 brachyrhizob
18	391	14.0	543	1 HPSE1_ARATH	Q9ff10 arabidopsis
19	372	13.3	527	2 Q9LRC8_SCUBA	Q9lrc8 scutellaria
20	362	12.9	526	2 Q5SNA6_ORYSA	Q5sna6 oryza sativ
21	361.5	12.9	541	2 Q69I16_ORYSA	Q69i16 oryza sativ
22	357.5	12.8	539	2 Q2QN56_ORYSA	Q2qn56 oryza sativ
23	355.5	12.7	536	1 HPSE3_ARATH	Q9fzpl arabidopsis
24	346.5	12.4	529	2 Q6ZJE2_ORYSA	Q6zje2 oryza sativ
25	346.5	12.4	537	2 Q70YJ3_HORVU	Q70yj3 hordeum vul
26	335	12.0	401	2 Q30324_ARATH	Q30324 arabidopsis
27	335.5	12.0	516	2 Q447R5_SOLUS	Q447r5 solibacter
28	310.5	11.1	539	1 HPSE2_ARATH	Q8l608 arabidopsis
29	309.5	11.1	537	2 Q43S03_SOLUS	Q43s03 solibacter
30	275.5	9.8	506	2 Q37Q70_SPHAR	Q37q70 novosphingo
31	242	8.7	382	2 Q3E8P7_ARATH	Q3e8p7 arabidopsis

RESULT 1					
ID	HPSE_MOUSE	STANDARD;	PRT;	535 AA.	
AC	Q6Ygz1, Q8K3K3;				
DT	11-OCT-2005, integrated into UniProtKB/Swiss-Prot.				
DT	11-OCT-2005, sequence version 2.				
DT	07-MAR-2006, entry version 13.				
DE	Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) (Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa subunit).				
GN	Name=Hpse; Synonyms=Hpa;				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;				
OC	Muridea; Muridae; Murinae; Mus.				
NCBI_TaxID	10090;				
[1]					
RN	NUCLEOTIDE SEQUENCE [MRNA].				
RC	STRAIN=SUJ/J; TISSUE=Spleen;				
RX	MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;				
RA	Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J., Parish C.R.;				
RT	"Cloning of mammalian heparanase, an important enzyme in tumor invasion and metastasis";				
RL	Nat. Med. 5:803-809(1999).				
[2]					
RN	NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 28-57 AND 150-179, GLYCOSYLATION, BIOPHYSICO-CHEMICAL PROPERTIES, ENZYME REGULATION, AND SUBUNITS.				
RC	STRAIN=FVB; TISSUE=Embryo;				
RX	MEDLINE=22350326; PubMed=12460766; DOI=10.1016/S1046-5928(02)00558-2;				
RA	Miao H.-Q., Navarro E., Patel S., Sargent D., Koo H., Wan H., Plata A., Zhou Q., Ludwig D., Bohlen P., Kussie P.;				
RT	"Cloning, expression, and purification of mouse heparanase.";				
RL	Protein Expr. Purif. 26:425-431(2002).				
[3]					
RN	NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.				
RP	MEDLINE=22841152; PubMed=12837765; DOI=10.1074/jbc.M300925200;				
RX	Gong F., Jemth P., Galvis M.L.E., Vlodavsky I., Horner A., Lindahl U., Li J.-P.;				
RA	"Processing of macromolecular heparin by heparanase.";				
RT	J. Biol. Chem. 278:35152-35158(2003).				
[4]					
RN	NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].				
RP	STRAIN=C57BL/6J, and NOD; TISSUE=Thymus;				
RX	PubMed=16141072; DOI=10.1126/science.1112014;				
RA	Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N., Oyama R., Ravasi S., Lenhard B., Wells C., Kodzius R., Shimokawa K., Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M., Davis M.J., Wilming L.G., Aidinis V., Allen J.E., Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L., Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M., Chiu K.P., Choudhary V., Christoffels B.P., de Bono B., Della Gatta G., Crowe M.L., Dalia E., Dalrymple B.P., de Bono B., Della Gatta G., di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,				

32	153	5.5	510	2	Q2U0T3 ASPOR	Q2u0t3 aspergillus
33	151	5.4	1128	2	Q5T65 ANOGA	Q5t65 anopheles g
34	138	4.9	559	2	Q7SFB0 NEUCR	Q7sfb0 neotospora
35	135.5	4.8	935	2	Q9VE79 DROME	Q9ve79 drosophila
36	133	4.8	670	2	Q3JTG0 BURP1	Q3jtg0 burkholderi
37	132	4.7	463	2	Q63T97 BURPS	Q63t97 burkholderi
38	130	4.6	1053	2	P77865 FIBSU	P77865 fibrobacter
39	129	4.6	1053	2	P71329 FIBSU	P71329 fibrobacter
40	124.5	4.5	1303	2	Q2ZH58 CALSA	Q2zh58 caldicellul
41	122.5	4.4	628	2	Q2ZEV8 CALSA	Q2zev8 caldicellul
42	118.5	4.2	536	2	Q2UDS9 ASPOR	Q2uds9 aspergillus
43	118.5	4.2	721	2	Q3XV07 9PROT	Q3xv07 magnetococ
44	118	4.2	2638	2	Q551W7 CRYNE	Q551w7 cryptococcu
45	118	4.2	2638	2	Q5KCT5 CRYNE	Q5ket5 cryptococcu







DT	11-OCT-2005, integrated into UniProtKB/Swiss-Prot.	FT	ACT_SITE	218	218	Proton donor (Potential).
DT	05-JUL-2004, sequence version 1.	FT	ACT_SITE	336	336	Nucleophile (Potential).
DT	07-MAR-2006, entry version 11.	FT	CARBOHYD	155	155	N-linked (GlcNAc. . .) (By similarity).
DE	Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) [Contains:	FT	CARBOHYD	193	193	N-linked (GlcNAc. . .) (By similarity).
DE	Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].	FT	CARBOHYD	210	210	N-linked (GlcNAc. . .) (By similarity).
GN	Name=Hpe; Synonyms=Hep;	FT	CARBOHYD	452	452	N-linked (GlcNAc. . .) (By similarity).
GN	Rattus norvegicus (Rat);	FT	CONFLICT	15	15	G -> R (in Ref. 2).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	FT	CONFLICT	227	227	H -> Q (in Ref. 2).
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;	FT	CONFLICT	350	350	D -> N (in Ref. 2).
OC	Muroidea; Muridae; Murinae; Rattus.	SQ	SEQUENCE	536 AA; 60480 MW; C434E04CF536EA4D CRC64;		
OX	NCBI_TaxID=10116;					
RP	NUCLEOTIDE SEQUENCE [MRNA].					
RC	TISSUE=Placenta;					
RC	MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;					
RA	Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,					
RA	Parish C.R.;					
RT	"Cloning of mammalian heparanase, an important enzyme in tumor					
RT	invasion and metastasis.";					
RL	Nat. Med. 5:803-809(1999).					
RL	[2]					
RP	NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.					
RP	MEDLINE=22194309; PubMed=12071130; DOI=10.1074/jbc.M203282200;					
RA	Podyma-Inoue K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;					
RA	"Characterization of heparanase from a rat parathyroid cell line.";					
RT	J. Biol. Chem. 277:32459-32465(2002).					
CC	-!- FUNCTION: Endoglycosidase which is a cell surface and					
CC	extracellular matrix-degrading enzyme. Cleaves heparan sulfate					
CC	proteoglycans (HSPGs) into heparan sulfate side chains and core					
CC	proteoglycans. Also implicated in the extravasation of leukocytes					
CC	and tumor cell lines. Contributes to metastasis and angiogenesis					
CC	(By similarity).					
CC	-!- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower					
CC	extent, by heparin and sulfamin (By similarity). Activated by					
CC	calcium and magnesium. Inhibited by EDTA.					
CC	-!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa					
CC	subunits, the proteolytic products (By similarity).					
CC	-!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.					
CC	Secreted, internalised and transferred to late endosomes/lysosomes					
CC	as a proheparanase. In lysosomes, it is processed into the active					
CC	form, the heparanase. The uptake or internalisation of					
CC	proheparanase is mediated by HSPGs. Heparin appears to be a					
CC	competitor and retain proheparanase in the extracellular medium					
CC	(By similarity).					
CC	-!- PTM: Proteolytically processed. The cleavage of the 65 kDa form					
CC	leads to the generation of a linker peptide, 8 kDa and 50 kDa					
CC	product. The active form, the 8/50 kDa heterodimer, is resistant					
CC	to degradation. Complete removal of the linker peptide appears to					
CC	be a prerequisite to the complete activation of the enzyme (By					
CC	similarity).					
CC	-!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears					
CC	to be essential for its solubility (By similarity).					
CC	-!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.					
CC	-----					
CC	Copyrighted by the UniProt Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a>					
CC	Distributed under the Creative Commons Attribution-NoDerivs License					
CC	-----					
DR	EMBL; AF359508; AAQ15189.1; -; mRNA.					
DR	EMBL; AF184967; AAF04563.1; -; mRNA.					
DR	RGD; 61969; Hpep.					
DR	InterPro; IPR005199; Glyco_hydro_79_N.					
DR	Pfam; PF03662; Glyco_hydro_79n; 1.					
KW	Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;					
FT	SIGNAL.					
FT	CHAIN 1 28 By similarity.					
FT	CHAIN 29 102 Heparanase 8 kDa subunit.					
FT						
FT	PROPEP 103 150 Linker peptide (By similarity).					
FT						
FT	CHAIN 151 536 /Frid-PRO_0000042267.					
FT						
FT	CHAIN 151 155 Heparanase 50 kDa subunit.					
FT	REGION 151 155 /Frid-PRO_0000042268.					
FT						
FT	REGION 263 273 Heparin/HS-binding (By similarity).					
FT						



```

RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
  cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
RC -----
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
  cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
  Distributed under the Creative Commons Attribution-NoDerivs License
  CC -----
DR EMBL: AM085492; CAJ30019.1; -: mRNA.
SQ SEQUENCE 574 AA; 64459 MW; 9F1D19DCBADD99DE CRC64;

Query Match      85.4%; Score 2387.5; DB 2; Length 574;
Best Local Similarity 84.9%; Pred. No. 1.4e-171;
Matches 455; Conservative 34; Mismatches 44; Indels 3; Gaps 2;

QY 1 MLRL-LLLLWGLGALAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPFLSITIDASLAT 59
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
41 MLRLSLLLWGLGSPVLCILA--AAQEDVVELEFSTQRPRLHLYSPFLSITIDANLAT 98

QY 60 DPRFLTFLGSPRLRALARGLSPAYLRFGGTKTDFLI FDPDKPTSEERSYWKSVQVNHDC 119
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
99 DPRFLTFLGSPKRALARGLSPAYLRFGGTKTDFLI FDPDKPEPSHEERSYWKSVQVNHDC 158

QY 120 RSEPVSAVLKQVLEWPFQELLRLREYQKEFKNSTYSRSSVDMLYSPAKCSGLDLIFG 179
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
159 RSGAIPAVVVRRLQVLEWPFQELLRLREYQKEFKNSTYSRSSVDMLYTPARCSGLDLIFG 218

QY 180 LNALLRTPDLRWSSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHLIDGLQGEDFV 239
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
219 LNALLRTADFRWSSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHLISIDGLQGEDYI 278

QY 240 ELHKLQRSFQNAKLYGPDIGPRGKTVKLLRSFLKAGGEVIDSLTWHHYLLNGRIATK 299
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
279 ELKLLRKSTLKNVLYGPDVGQPRGKTVKLLRSFLKAGGEVIDSVTWHHYLLNGRIATK 338

QY 300 EDFLSSDALDTFILSVQKILKVTKETITPKKVMWLGTSAYGGAPLLSNTFAAGFMWLD 359
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
399 KLGLSAQMGIEVVRQVFFGAGNYHLVDENFEPLPDYMLSLFKLVGPRVLLSRVKGPD 458

QY 420 RSKLRVYLHCTNVHPRYQEGDLTLVNLHNVTGHLKVPPLPRKVPDVTYLLKPSGPDG 479
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
459 RSKLRVYLHCTNINHPRYQEGDLTLVNLHNVTGHLKVPPLPRKVPDVKYLVKPLPGG 518

QY 480 LLSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAACI 535
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
519 LLSKSVQLNGQALKWDDQTLPALTEKPLRPGSSLSGLPAFSYGFFVIRNAKVAACL 574

RESULT 6
Q333X6 SPAJD
ID Q333X6 SPAJD PRELIMINARY; PRT; 574 AA.
AC Q333X6
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax judaei (Blind subterranean mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=334510;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
  cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).

```

```

RT cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
RC -----
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
  cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
  Distributed under the Creative Commons Attribution-NoDerivs License
  CC -----
DR EMBL: AM085493; CAJ30020.1; -: mRNA.
SQ SEQUENCE 574 AA; 64515 MW; 3ABEB13F07451684 CRC64;

Query Match      85.1%; Score 2379.5; DB 2; Length 574;
Best Local Similarity 84.7%; Pred. No. 5.5e-171;
Matches 454; Conservative 34; Mismatches 45; Indels 3; Gaps 2;

QY 1 MLRL-LLLLWGLGALAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPFLSITIDASLAT 59
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
41 MLRLSLLLWGLGSPVLCILA--AAQEDVVELEFSTQRPRLHLYSPFLSITIDANLAT 98

QY 60 DPRFLTFLGSPRLRALARGLSPAYLRFGGTKTDFLI FDPDKPTSEERSYWKSVQVNHDC 119
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
99 DPRFLTFLGSPKRALARGLSPAYLRFGGTKTDFLI FDPDKPEPSHEERSYWKSVQVNHDC 158

QY 120 RSEPVSAVLKQVLEWPFQELLRLREYQKEFKNSTYSRSSVDMLYSPAKCSGLDLIFG 179
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
159 RSGAIPAVVVRRLQVLEWPFQELLRLREYQKEFKNSTYSRSSVDMLYTPARCSGLDLIFG 218

QY 180 LNALLRTPDLRWSSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHLIDGLQGEDFV 239
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
219 LNALLRTADFRWSSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHLISIDGLQGEDYI 278

QY 240 ELHKLQRSFQNAKLYGPDIGPRGKTVKLLRSFLKAGGEVIDSLTWHHYLLNGRIATK 299
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
279 ELKLLRKSTLKNVLYGPDVGQPRGKTVKLLRSFLKAGGEVIDSVTWHHYLLNGRIATK 338

QY 300 EDFLSSDALDTFILSVQKILKVTKETITPKKVMWLGTSAYGGAPLLSNTFAAGFMWLD 359
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
399 KLGLSAQMGIEVVRQVFFGAGNYHLVDENFEPLPDYMLSLFKLVGPRVLLSRVKGPD 419

QY 420 RSKLRVYLHCTNVHPRYQEGDLTLVNLHNVTGHLKVPPLPRKVPDVTYLLKPSGPDG 479
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
459 RSKLRVYLHCTNINHPRYQEGDLTLVNLHNVTGHLKVPPLPRKVPDVKYLVKPLPGG 518

QY 480 LLSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAACI 535
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
519 LLSKSVQLNGQALKWDDQTLPALTEKPLRPGSSLSGLPAFSYGFFVIRNAKVAACL 574

RESULT 7
Q333X5 SPAJD
ID Q333X5 SPAJD PRELIMINARY; PRT; 558 AA.
AC Q333X5
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax judaei (Blind subterranean mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=334510;
RN [1]
RP NUCLEOTIDE SEQUENCE.

```



Db 74 FTELGSKLRTLARGAPAYLREGNGKDFLLFDPKKEPAFEERSYWLQSNQDICKSGS 133  
Qy 124 VSAVLRKQVQPPQELLRLRQYQKEPKNSYRSVDMLYSFACSGLDLIFGNAL 183  
Db 134 IPSDVBEKLRLEWPFQEQVLLREQYQKFTNSTYRSVDMLYTFACSGLNLIIFGNAL 193  
Qy 184 LRTPTDLRWSSNAQLLDYCSSKYNLSWELGNEPNSFWKKAHLIDGLQLOLGEDFVHLK 243  
Db 194 LRTTDMHWDSSNAQLLDYCSSKYNLSWELGNEPNSFORKAGIFINGRQLOGDFIBFRK 253  
Qy 244 LLORSFQNAKLXGPDIGQPRGTGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKEDFL 303  
Db 254 LLKSFAFKNAKLXGPDIGQPRGTGKTVKLLRSFLKAGGEVIDSVTWHYYVNGRIATKEDFL 313  
Qy 304 SSDALFTFSLVQKILKVTKEITPGKKVWLGETSSAYGGAPLNSNTFAAGFWMLDKLGL 363  
Db 314 NPDILTFFISSVQKTRIVEKIRPKVWLGETSSAFGGGAPLNSNTFAAGFWMLDKLGL 373  
Qy 364 SAQMGIEVVMRQVFFGAGNHYLVDEPEPLDYWLSLLFKKLVGPRVLLSRVKGPDRLSKL 423  
Db 374 SARMGIEVVMRQVLFAGNHYLVDEPEPLDYWLSLLFKKLVGPRVLLSRVKGPDRLSKF 433  
Qy 424 RYVLHCTNVYHPRYQSGDLTYVLNLHNVTKHLKVPPLERKPVDTVLLKPSGPDGLLSK 483  
Db 434 RYVLHCTNVKHPRYKSGDLTYVLNLHNVTKHLKVPPLERKPVDTVLLKPSGPDGLLSK 493  
Qy 484 SVQLNQILKMWDEQTLPALTEKPLPAGSALSIPAFSGYGFVFNRAKIAACI 535  
Db 494 SVQLNQILKMWDEQTLPALTEKPLPAGSLSGMPFSGYGFVFNRAKIAACI 545

## RESULT 9

HPSE HUMAN  
ID HPSE HUMAN STANDARD; PRT; 543 AA.  
AC Q9Y2E1; Q53G5; Q9UL39;  
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
DT 01-NOV-1999, sequence version 1.  
DT 07-FEB-2006, entry version 27.  
DE Heparanase precursor [EC 3.2.-.-] (Heparanase-1) (Hpal) (Endo-glucuronidase) [Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].  
DE subunit.  
GN Name=HPSE; Synonyms=HEP, HPA, HPA1, HPR1, HPSE1, HSE1;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.  
RC TISSUE=Placenta;  
RX MEDLINE=99335379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;  
RA Kusile P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C., Seddon A.P., Giorgio N.A., Bohlen P.;  
RT "Cloning and functional expression of a human heparanase gene.";  
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND PROTEIN SEQUENCE OF 158-168; 326-337 AND 447-491.  
RC TISSUE=Embryonic fibroblast;  
RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;  
RA Toyoshima M., Nakajima M.;  
RT "Human heparanase. Purification, characterization, cloning, and expression.";  
RL J. Biol. Chem. 274:24153-24160(1999).  
RN [3]  
RP NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.  
RX PubMed=10395325; DOI=10.1038/10518;  
RA Vlodavsky I., Friedmann Y., Elkin M., Aingorn H., Atzmon R., Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I., Spector L., Pecker I.;  
RT "Mammalian heparanase: gene cloning, expression and function in tumor progression and metastasis.";

Nat. Med. 5:793-802(1999).  
RN [4]  
RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE OF 158-174; 263-272; 326-337; 433-436; 438-443; 466-468 AND 478-483.  
RC TISSUE=Placenta;  
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;  
RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J., Parish C.R.;  
RT "Cloning of mammalian heparanase, an important enzyme in tumor invasion and metastasis.";  
RL Nat. Med. 5:803-809(1999).  
RN [5]  
RP NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICOCHEMICAL PROPERTIES, AND TISSUE SPECIFICITY.  
RC TISSUE=Placenta;  
RX MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;  
RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;  
RT "Heparanase expression in invasive trophoblasts and acute vascular damage.";  
RL Glycobiology 10:467-475(2000).  
RN [6]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.  
RX PubMed=11547900; DOI=10.1023/A:1011375624902;  
RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O., Aviv A., Elkin M., Pecker I., Peretz T., Vlodavsky I.;  
RT "Molecular properties and involvement of heparanase in cancer progression and mammary gland morphogenesis.";  
RL J. Mammary Gland Biol. Neoplasia 6:311-322(2001).  
RN [7]  
RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163, SUBUNITS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.  
RC TISSUE=Placenta;  
RX PubMed=12713442; DOI=10.1042/BJ20030318;  
RA McKenzie E., Young K., Hircok M., Bennett J., Bhawan M., Felix R., Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S., Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M., Parekh R., Stubberfield C.;  
RT "Biochemical characterization of the active heterodimer form of human heparanase (Hpal) protein expressed in insect cells.";  
RL Biochem. J. 373:423-435(2003).  
RN [8]  
RP NUCLEOTIDE SEQUENCE [MRNA].  
RA Pinhal M.A., Semedo P.;  
RT "Cloned heparanase from MCF-7 cells.";  
RN Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.  
RN [9]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC TISSUE=Small intestine;  
RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y., Tanaka A., Yokoyama S.;  
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.  
RN [10]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC TISSUE=Pancreas;  
RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Beak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";  
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[11]  
RP MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.  
RX PubMed=11123890; DOI=10.1021/bi002080p;  
RA Hulett M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,  
RA Gready J.E., Farish C.R.;  
RT "Identification of active-site residues of the pro-metastatic  
RT endoglycosidase heparanase.";  
RL Biochemistry 39:15659-15667(2000).  
[12]  
RN N-GLYCOSYLATION, AND MUTAGENESIS OF ASN-162; ASN-178; ASN-200;  
RP ASN-217; ASN-238 AND ASN-459.  
RX PubMed=14573609; DOI=10.1074/jbc.M300541200;  
RA Simizu S., Ishida K., Wierzbza M.K., Osada H.;  
RT "Secretion of heparanase protein is regulated by glycosylation in  
RT human tumor cell lines.";  
RL J. Biol. Chem. 279:2697-2703(2004).  
[13]  
RN SUBCELLULAR LOCATION.  
RP PubMed=15292202; DOI=10.1074/jbc.M402131200;  
RA Gingis-Velitski S., Zetser A., Kaplan V., Ben-Zaken O., Cohen E.,  
RA Levy-Adam F., Bashenko Y., Flugelman M.Y., Vlodaysky I., Ilan N.;  
RT "Heparanase uptake is mediated by cell membrane heparan sulfate  
RT proteoglycans.";  
RL J. Biol. Chem. 279:44084-44092(2004).  
[14]  
RN BIOPHYSICO-CHEMICAL PROPERTIES, PROCESSING, AND SUBCELLULAR LOCATION.  
RX PubMed=15848168; DOI=10.1016/j.febslet.2005.03.030;  
RA Cohen E., Atzmon R., Vlodaysky I., Ilan N.;  
RT "Heparanase processing by lysosomal/endosomal protein preparation.";  
RL FEBS Lett. 579:2334-2338(2005).  
[15]  
RN SUBCELLULAR LOCATION, PROCESSING, AND MUTAGENESIS OF TYR-156.  
RX PubMed=15659389; DOI=10.1074/jbc.M413370200;  
RA Abboud-Jarroos G., Rangini-Guetta Z., Aingorn H., Atzmon R.,  
RA Elgavish S., Peretz T., Vlodaysky I.;  
RT "Site-directed mutagenesis, proteolytic cleavage, and activation of  
RT human proheparanase.";  
RL J. Biol. Chem. 280:13568-13575(2005).  
[16]  
RN DOMAINS, AND MUTAGENESIS OF LYS-158 AND LYS-161.  
RX PubMed=15760902; DOI=10.1074/jbc.M414546200;  
RA Levy-Adam F., Abboud-Jarroos G., Guerrini M., Beccati D.,  
RA Vlodaysky I., Ilan N.;  
RT "Identification and characterization of heparin/heparan sulfate  
RT binding domains of the endoglycosidase heparanase.";  
RL J. Biol. Chem. 280:20457-20466(2005).  
[17]  
RN VARIANT SER-260.  
RX PubMed=15334672;  
RA Chen X.P., Liu Y.B., Rui J., Peng S.Y., Peng C.H., Zhou Z.Y.,  
RA Shi L.H., Shen H.W., Xu B.;  
RT "Heparanase mRNA expression and point mutation in hepatocellular  
RT carcinoma.";  
RL World J. Gastroenterol. 10:2795-2799(2004).  
CC -!- FUNCTION: Endoglycosidase which is a cell surface and  
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
CC proteoglycans. Also implicated in the extravasation of leukocytes  
CC and tumor cell lines. Due to its contribution to metastasis and  
CC angiogenesis, it is considered to be a potential target for anti-  
CC cancer therapies.  
CC -!- ENZYME REGULATION: Inhibited by EDTA, laminarin sulfate and, to a  
CC lower extent, by heparin and sulfamin and activated by calcium and  
CC magnesium (By similarity).  
CC -!- BIOPHYSICO-CHEMICAL PROPERTIES:  
CC pH dependence:  
CC Optimum pH is 4-6;  
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
CC subunits, the proteolytic products.  
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.  
CC Secreted, internalised and transferred to late endosomes/lysosomes

CC as a proheparanase. In lysosomes, it is processed into the active  
CC form, the heparanase. The uptake or internalisation of  
CC proheparanase is mediated by HSPGs. Heparin appears to be a  
CC competitor and retain proheparanase in the extracellular medium.  
CC -!- TISSUE SPECIFICITY: Highly expressed in placenta and spleen and  
CC weakly expressed in lymph node, thymus, peripheral blood  
CC leukocytes, bone marrow, endothelial cells, fetal liver and tumor  
CC tissues.  
CC -!- PTM: Proteolytically processed. The cleavage of the 65 kDa form  
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa  
CC product. The active form, the 8/50 kDa heterodimer, is resistant  
CC to degradation. Complete removal of the linker peptide appears to  
CC be a prerequisite to the complete activation of the enzyme.  
CC -!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears  
CC to be essential for its solubility.

Query Match 76.8%; Score 2149; DB 1; Length 543;  
Best Local Similarity 76.6%; Pred. No. 1.4e-153;  
Matches 407; Conservative 50; Mismatches 74; Indels 0; Gaps 0;  
QY 5 LLLWMLGALGALGAPAGTAPTDVVDLEFVYKRPRLSRVSPSFLSITIDASLATDPRFL 64  
DB 13 LMLLLGLPLSPGALPRPAQADVDLDFQEPHLVSPSFLSITIDANLATDPRFL 72  
QY 65 TFGSPRLRALARGLSPAYLRFGTKTDFLI FPDPKPTSEERSYKQSVNNDICRSEPV 124  
DB 73 ILLGSPKRLTARGLSPAYLRFGTKTDFLI FDKPKSTFEERSYQSQVNQDICKYGI 132  
QY 125 SAAVLRKLVQWEPFOELLRLREQVQKEFKNSTYSRSSVDMLYSFAKCSGLDLIFGLNALL 184  
DB 133 PPVEEKLRLWEVPYQEQLLREHYQKKFNSTYSRSSVDVLYTTFANCGLDLIFGLNALL 192  
QY 185 RTPDLRWNSNAQLLLDYCSSKYNISWELGNENPSFWKKAHILIDGLQAGEDFVELHKL 244  
DB 193 RTADLQWNSNAQLLLDYCSSKYNISWELGNENPSFKKADIFINGSLQGEDFIQLHKL 252  
QY 245 LQSAFQNAKLYGPDIGQPRKTVKLLRSFLKAGEVIDSLTWHYLYNGRIATKEDFLS 304  
DB 253 LRKSTFNKAKLYGPDVGQPRKTKAKMLKSLKAGEVIDSLTWHYLYNGRTATREDFLN 312  
QY 305 SDALDTFILSVQKLLKVTKEITPGKKVWLGTSAYGGAPLLSNTFAAGFWMLDKGLS 364  
DB 313 PDVLDIFISSVQVQVVESTRPGKKVWLGTSAYGGAPLLSNTFAAGFWMLDKGLS 372  
QY 365 AOMGIEVWQVFFGAGNYHLVDENFPLPDYWLSSLFKLVGPRVLLSRVKGPDRSKLR 424  
DB 373 ARMGIEVWQVFFGAGNYHLVDENFPLPDYWLSSLFKLVGTVKVLMAVQGSKRKLR 432  
QY 425 VYLHCTNVVHPRYQEGDLTLVNLNHNVTXKLVKPPPLFRKPVDTYLLKPSGPDGLLSKS 484  
DB 433 VYLHCTNDNPRYKEGDLTLVNLNHNVTXKLVKPPPNQVQDKVLLRPLGPHGLLSKS 492  
QY 485 VOLNGQLKQWDEQTLTPALTEKELPAGSALSPLAFSGFFVIRNAKTAACI 535  
DB 493 VOLNGLTLKMWDDQTLPLMEKPLRPGSLGSLPAGSFFVIRNAKVAACI 543

RESULT 10  
HPSE CHICK  
ID HPSE\_CHICK STANDARD; PRT; 523 AA.  
AC Q30YK5;  
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
DT 01-DEC-2001, sequence version 1.  
DT 07-FEB-2006, entry version 12.  
DE Heparanase precursor (EC 3.2.-.-).  
GN Name=HPSE; Synonyms=HPA;  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OC NCBI TaxID=9031;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA].



RX MEDLINE=21369959; PubMed=11387326; DOI=10.1074/jbc.M102462200;  
RA Goldshmidt O., Zcharia E., Angioni E., Guatta-Rangini Z., Atzmon R.,  
RA Michal I., Fecker I., Mitrani E., Vlodavsky I.;  
RT "Expression pattern and secretion of human and chicken heparanase are  
determined by their signal peptide sequence.";  
RL J. Biol. Chem. 276:29178-29187 (2001).  
CC -I- FUNCTION: Endoglycosidase which is a cell surface and  
extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
proteoglycans (HSPGs) into heparan sulfate side chains and core  
proteoglycans (By similarity).  
CC -I- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted  
(By similarity).  
CC -I- PTM: N-glycosylated (By similarity).  
CC -I- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; AY037007; AAK82648.1; -; mRNA.  
DR Ensembl; ENSGALG0000011203; Gallus gallus.  
DR InterPro; IPR005199; Glyco\_hydro\_79\_N.  
DR Pfam; PF03662; Glyco\_hydro\_79n; 1.  
KW Glycoprotein; Hydrolase; Lysosome; Membrane; Signal.  
FT SIGNAL 1 18 Potential.  
FT CHAIN 19 523 Heparanase.  
FT  
FT REGION 137 141 Heparin/Hs-binding (By similarity).  
FT REGION 250 260 Heparin/Hs-binding (By similarity).  
FT ACT\_SITE 204 204 Proton donor (Potential).  
FT ACT\_SITE 323 323 Nucleophile (Potential).  
FT CARBOHYD 141 141 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 196 196 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 436 436 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 439 439 N-linked (GlcNAc...) (Potential).  
FT SEQUENCE 523 AA; 58386 MW; 8EB07B18C9BF891 CRC64;  
SQ  
Query Match 57.4%; Score 1605; DB 1; Length 523;  
Best Local Similarity 58.8%; Pred. No. 1.9e-112;  
Matches 315; Conservative 81; Mismatches 126; Indels 14; Gaps 4;  
QY 1 MLRLLLMLWGLGALQAQAPAGTAPDVTDDVDFYKRLRSPFLSTIDASLATD 60  
DB 1 MLVLLLLVL-----LNAVPPRTA-----ELQLGREPIGAVSPAFSLTLDASLARD 48  
QY 61 PRFLTEFLGSPRLALARGLPAYLRFSGTGTDTPLIFDPDKTEPSEERSYKWSQVNHICR 120  
DB 49 PRFVALLRHPKLHTLASGLSPGFURFGTSTDFLIFPNKDSWTWEKVLSEFQAK-DVCE 107  
QY 121 SEPVSAALVRKLQVWEPFQELLRLREQYKFEKNSTYSRSVDMLYSFAKSGLDLIFGL 180  
DB 108 AWPSPFAVVPKLLLTQWLPQEKLLLAHSHKXKHKTITITRSLDILHTFASSSGFLVFG 167  
QY 181 NALLRTPLDRLWNSNAQLLDYCSKGYNIISWELGNFSPFWKKAHLIDGLQIGEPFVE 240  
DB 168 NALLRRAGLQWDSNNAQLLYCYAQRGSYNIISWELGNFSPFWKKAHLIDGLQIGEPFVH 227  
QY 241 LHKLL-ORSFAFONAKLPGDIPGQPRGKTVKLLRFLKAGGEVIDSLTWHVYLLNGRTATK 299  
DB 228 LRQLLSQHPYRHAELYGLDVGQPRKHTQHLRFLKAGGEVIDSLTWHVYLLNGRTATK 287  
QY 300 EDFLSSDALDTFILSVQILKLVKTEITPGKKVWLGSETSSAYGGCAPLLSNTFAAGFWMLD 359  
DB 288 EDFLSPVLDSTATAHDLVIGVETATPGKKVWLGSETSSAYGGCAPLLSNTFAAGFWMLD 347  
QY 360 KLGLSAQMGIEVVRQVFFGAGNYHLVDENFEPLDYWLSSLFKLVGPRVLLSRVKGPD 419  
DB 348 KLGLAARGIDVVRQVFSFGAGSYHLVDAGFKPLDPDYWLSSLFKLVGTRVLQASVEQAD 407  
QY 420 RSKLRVYLHCTNVTHRYQEGDLFLYVNLHNVTKHLKVPPLPRKVPDVTLLKPSGPDG 479  
DB 408 ARRPRVYLHCTNPRHPKYREGDVTFLALNLSNVTQSLQPLQKLSKSDVQYLLPHGKDS 467  
QY 480 LLSKSVQLNGQILKMWDEQTLTPALTEKPLPAGSALSUPAFSYGFFVIRNAKIAACI 535

DB 468 ILSREVQLNGRLQWVDETLPALHEMALAPGSLGLPAFSYGFVIRNAKIAACI 523  
RESULT 11  
Q4SYF6\_TETNG PRELIMINARY; PRT; 533 AA.  
AC Q4SYF6;  
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
DT 19-JUL-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Chromosome undetermined SCAF12073, whole genome shotgun sequence.  
DE (Fragment).  
GN ORFNames=GSTENG00010356001;  
OS Tetraodon nigroviridis (Green puffer).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
OC Tetraodontidae; Tetraodontidae; Tetraodon.  
OX NCBI\_TaxID=99883;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15496914; DOI=10.1038/nature03025;  
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,  
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,  
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,  
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
RA Anthonard V., Jubin C., Castellani V., Katinka M., Vacherie B.,  
RA Bismont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,  
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Guzy J.,  
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,  
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,  
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
RA Wincker P., Lander E.S., Weissbach J., Roest Crollius H.;  
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals  
the early vertebrate proto-karyotype.";  
RL Nature 431:946-957 (2004).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RG Genoscope; Whitehead Institute Centre for Genome Research;  
RL Submitted (FEB-2004) to the EMBL/GenBank/DBSJ databases.  
CC -I- CAUTION: The sequence shown here is derived from an  
EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is  
preliminary data.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
DR EMBL; CAAR01012073; CAF94326.1; -; Genomic\_DNA.  
FT NON\_TER 1 533  
FT NON\_TER 533 533  
FT SEQUENCE 533 AA; 60100 MW; 9B00A7C8780100FF CRC64;  
SQ  
Query Match 46.1%; Score 1289; DB 2; Length 533;  
Best Local Similarity 48.2%; Pred. No. 1.5e-88;  
Matches 257; Conservative 94; Mismatches 144; Indels 38; Gaps 8;  
QY 41 LRSVSPFLSTIDASLATDPRFLTFLGSPRLALARGLSPAYLRFSGTGTDTPLIFDPDK 100  
DB 1 LRRVDPFLSVTIDASLAADERFMVLLSSPKVTILAKALTAPLRFPGGTRDFMVFAPHK 60  
QY 101 -EPTS--EERSYKWSQVNHICRSEPVSAALVRKLQVWEPFQELLRLREQYKFEKNSTY 157  
DB 61 NOPASGFSARELITSSNGQHCSEKAPPPWLERRLRTEWKKQVMLRNEBLQRKRVAF 120  
QY 158 SRSSVDMLYSFAKSGLDLIFGLNALLRTPDLRWNSSNAQLLDYCSKGYNIISWELGNE 217  
DB 121 TETTVQDLHAFANCSGLDLVFLGNALLRTADNRWNSSNARSLLRYCEARRYHMSWELGNE 180  
QY 218 PMSFWKKAHLIDGLQGEDFVELHKLQRSFAF-QNAKLYGPDIGOPRGTVKLLRSFLK 276  
DB 181 PMSYKKAHLIDGLQGEDFVELHKLQRSFAF-QNAKLYGPDIGOPRGTVKLLRSFLK 240







<p>[2]          RN NUCLEOTIDE SEQUENCE.          RP TISSUE=PCR rescued clones;          RG NIH MGC Project;          RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.          CC -----          CC Copyrighted by the UniProt Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a>          CC Distributed under the Creative Commons Attribution-NoDerivs License          CC -----          DR EMBL; BC112356; AA112357.1; -; mRNA.          SQ SEQUENCE 592 AA; 66610 MW; 9468981C2A74359F CRC64;</p>	<p>Query Match 41.8%; Score 1169.5; DB 2; Length 592;          Best Local Similarity 46.4%; Pred. No. 2e-79;          Matches 245; Conservative 80; Mismatches 176; Indels 27; Gaps 6</p>
<p>Qy 33 LEFTTKPLRSVSPFLSIITDASLATDPRFLTFLGSPRLRALARGCLSPAYLRFGCTKD 92          Db 63 LDVTTKQNPVRTNENFLSLQDPSIHD-GWLDFLSSKRLVTLARGLSPAFLRFGCKTD 121          Qy 93 FLIF-----DPQKPTSEERSYWKSVQNHDI CRSEPV-----SAAVLRLKQVEMP 137          Db 122 FLQFNLRNPAKSRGGPGDYILKNYEDDI VRSVDVALDKQCKCKTAQHPDVMLELQREKA 181          Qy 138 FQ-ELLILLREYQKFEKNSTYSRSQVDMLYSPAKCSGLDLIFGLNALLRTPDLRWNSSNA 196          Db 182 AQHMLVLLKEQFSNTYSNLLLTARSLDKLYNFADCSGLHLIFALNALRNPNNSSSSA 241          Qy 197 QLLLDYCSSKKYNIISWELGNENPSPFWKKAHLIDGLQIGEDFVELHKLQIR-SAFQNAKL 255          Db 242 LSLLEKYGASKKYNIISWELGNENPNYRTHWGRAVNSGQLGKDYIQLKSLQPIRIYSRSL 301          Qy 256 YGPDIGOPRGKTVKLLRSFLKAGGEVIDSLTWHVYLLNGRIATKDEDFSSDALDPFILSV 315          Db 302 YGPNIGRPRKNVIALLDGFMKVAGSTVDVAVTHQCHYIDGRVVKVMDFLKTRLLDTLSQI 361          Qy 316 QKILKVTKEITPGKKVWLGETSSAYGGGAPILLSNTFAAGFMWLDKGLGSAQGWIEVNMQR 375          Db 362 RKIQKVVNTYTPGKKIWLGVVTTSGGTTNNLSDYAAGFLWLTNLTGLMANQIGIDVIRH 421          Qy 376 VFFGAGNYHLVDENPEPLPDYWLSSLFFKLVGPRVLLSRVKGPD-----RSKLAVY 426          Db 422 SFFDHGYNHLVDQNFNLPDYLWLSLLYKRLGPKVLAVHVAGLQRKPRGPRVIRDKLRIY 481          Qy 427 LHCTNHPHYRQOEGDLYLVNLHNVTYHKLKVPPLPRKPPVDYTLTKPSGPDGLLSKSVQ 486          Db 482 AHCTNHHNNYVRGSIITLFIINLHRSRKIKLTGTIRDKLVHYYLLQYPQOEGLSKSVQ 541          Qy 487 LNGQILKMWDSQTLPALTEKPLPAGSALSPLAFSYGFFVIRNAKTAAC 534          Db 542 LNGQPLVMVDGTLPELKPRPLRAGRTLVIPEVTMGFFWKNVNALAC 589</p>	

RESULT 14  
Q4TB80 TETNG PRELIMINARY; PRT; 597 AA.  
ID Q4TB80 TETNG  
AC Q4TB80;  
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
DC 19-JUL-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Chromosome 17 SCAF7180, whole genome shotgun sequence. (Fragment).  
GN ORFNAME=GSTENG0003868001;  
OS Tetraodon nigroviridis (Green puffer).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
OC Tetraodontidae; Tetraodontidae; Tetraodon.  
OX NCBI TaxID=99983;  
OX

[1]  
NUCLEOTIDE SEQUENCE.  
RP  
RX  
RA  
RX  
RA  
RA  
RA  
MA  
MA  
MA

PubMed=15496914; DOI=10.1038/nature03025;  
Jallou O., Aubry J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,  
Maucelli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,

RA	Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA	Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA	Anchovard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA	Bismont C., Skalli Z., Cattolico L., Poulain J., De Barardinis V.,
RA	Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA	Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA	Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA	Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA	Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA	Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.,
RT	"Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT	the early vertebrate proto-karyotype.";
RL	Nature 431:946-957(2004).
RP	[2]
RP	NUCLEOTIDE SEQUENCE.
RG	Genoscope; Whitehead Institute Centre for Genome Research;
RL	Submitted (FEB-2004) to the ENBL/GenBank/DBJ databases.
CC	-!- CAUTION: The sequence shown here is derived from an
CC	ENBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC	preliminary data.
CC	-----
CC	Copyrighted by the UniProt Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a>
CC	Distributed under the Creative Commons Attribution-NoDerivs License
CC	-----
EMBL	EMBL; CAEA01007180; CAF89852.1; -; Genomic_DNA.
DR	NON TER 597
FT	SEQUENCE 597 AA: 67127 MW: 83B46AC5B727A8FE CRC64:
SQ	

Query Match	37.3%	Score 1042.5	DB 2	Length 597
Best Local Similarity	42.6%	Pred. No. 8e-70		
Matches	230	Conservative 84	Mismatches 187	Indels 39
Gaps	10			
Qy	33	LEFVTKPLRSVSPFLSIITIDASLATDPRFLTFLGSPRLALARGLSPAYLRFQGT	KTD	92
Db	58	LDVNTRSPIRVLNDFLSQLDPSIHKD-GWLDLFLSSGRLVTLARGLSPAFLRFQGKRTD	116	
Qy	93	FLIPDPKE-----PTSEERSVWKSVQNHDI-----CRSEPVSAAVLRLKQVWP	137	
Db	117	FLOFTNQKLUAKFRGPGPDYVLYKNYEDDIIRSDIALDKQCKGLASHPDMMLELQREKAA	176	
Qy	138	FOEILLREOVQKEFKNST-----YSRS-----SDVMWYSFAKCSGLDLIFGLNALRT	PD	189
Db	177	STQVLVLKEQLSNYSNITLTGLESHRIARSLSDKLYNFADCAGLHILIGLNALHRNPDH	236	
Qy	190	RWNSSNAQLLDYCSKGYNISWELGNEPNSFWKKAHILIDQLGIDGFVELHLKLQR-S	248	
Db	237	SWNTSSTLSLLKYSGAKKYNISWELGNEPNAYRSMVGHAVNSSQLAQDYTKLRTLQSVR	296	
Qy	249	AFQNAKLYGDDIGOPROKTVKLLRSFLKAGEVIDSLTWHYYLYNGRTATKEDFLUSSDAL	308	
Db	297	YYSRAOILYGNAGRPKNALILLDFEMKTVGTVDVATWQHYMDGR1KKYVEDFLKTRLL	356	

Qy	309	DTFILSVOKILKVTKETPGKKVWLGSETSSAYGGAPLLSNTNTPAAGFMWLDKGLGSAQMG	368
Db	357	DTLTELQLSKVTVVNTHTPGKKWLGGLGPAWTCGMSNLSDTFAAGFLVWNTLGNAAQMG	416
Qy	369	IEVVMR-----QVFGAGNYHLVDENPEP-LPDYWLSSLFKLVGVRVLLSRVKG----	419
Db	417	IDVVLRRQAQVEHTNKQSVALFLQMFVPSFDYVQSLVFKLVGPKVLAVRAGLQRKQP	476
Qy	420	-----REKLRYVLCTNMYHPRYCEGDLTLVLNLHNVTXKLVKVPPLFRKPVDPVLYLKP	474
Db	477	PGRVIRDKLRIYACTSYSNHNYVYRGSIYIILNHSRKKIKLAGTLRNNIHVHQLQLP	536
Qy	475	SGPDGLLSKSVQLNGQILKMVDEQTLPALTEKPLPAGSALSPLAFSYGFFVIRNAKIAAC	534
Db	537	YGADGLRAKHVQNGEKLIMADNETFPELKPRTLRAGRTIAMPMTIIFYVYIKNLINAVAC	596

RESULT 15  
Q4TGC8\_TETNG  
ID Q4TGC8\_TETNG PRELIMINARY; PRT; 255 AA.  
AC Q4TGC8;



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:10:07 ; Search time 17.8121 seconds  
(without alignments)  
2889.939 Million cell updates/sec

Title: US-10-645-659A-2  
Perfect score: 2797  
Sequence: 1 MLRLLLLMGLGALQAQA.....LPAPSYGFFVIRNAKIAACI 535

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	895.5	32.0	480	JC7506	heparanase protein
2	391	14.0	521	T45608	hypothetical prote
3	161.5	5.8	190	T01953	hypothetical prote
4	117	4.2	1331	A48954	mannan endo-1,4-be
5	114.5	4.1	1573	T50113	3-dehydroquinatase
6	113.5	4.1	837	A31842	endo-1,4-beta-xyla
7	111.5	4.0	599	T575363	hypothetical prote
8	110	3.9	575	T45668	hypothetical prote
9	109.5	3.9	2165	RRN2A2	genome polypotein
10	105.5	3.8	897	G02529	dynein heavy chain
11	105.5	3.8	464	A38905	dynein heavy chain
12	104.5	3.7	587	S36231	beta-fructofuranos
13	104	3.7	484	G72395	alpha-L-arabinofur
14	103.5	3.7	960	T37916	probable heterochr
15	103	3.7	464	C40630	GDP-mannose pyroph
16	103	3.7	500	D87541	beta-xylosidase [i
17	103	3.7	557	ODNC1	cytochrome-c oxida
18	102.5	3.7	505	G86391	hypothetical prote
19	102	3.6	596	T04506	ATP-dependent nucl
20	102	3.6	1180	A13504	cytochrome-c oxida
21	101.5	3.6	523	T11317	formin-binding pro
22	101.5	3.6	844	T52396	DNA topoisomerase
23	101.5	3.6	1462	T06819	hypothetical prote
24	100.5	3.6	335	T05268	hypothetical prote
25	100.5	3.6	356	D85406	hypothetical prote
26	100	3.6	714	CIUH	calpain (EC 3.4.22
27	99.5	3.6	1392	1 YGBVAD	L-aminoadipate-sem
28	99	3.5	709	A35364	carcinoembryonic a
29	99	3.5	824	T10615	hypothetical prote

30	98.5	3.5	772	2	S62481	hypothetical prote
31	98.5	3.5	4436	2	E71086	hypothetical prote
32	98.5	3.5	7962	2	I38346	elastic titin-hu
33	98	3.5	541	2	A48327	cytochrome-c oxida
34	98	3.5	1288	2	T09908	hypothetical prote
35	97.5	3.5	496	2	T04020	hypothetical prote
36	97.5	3.5	1708	2	A05205	hypothetical prote
37	97	3.5	511	2	S61166	probable membrane
38	97	3.5	796	2	D97065	transketolase [imp
39	97	3.5	1482	2	S13495	pregnancy zone pro
40	96.5	3.5	361	2	A43510	basic membrane pro
41	96.5	3.5	411	2	S74760	hypothetical prote
42	96.5	3.5	641	2	A42019	tyrosine-tRNA liga
43	96	3.4	439	2	H69783	conserved hypotet
44	96	3.4	725	1	P21VBS	RNA-directed RNA p
45	95.5	3.4	263	2	AF2534	hypothetical prote

ALIGNMENTS

RESULT 1

JC7506

heparanase protein 2a - human  
C:Species: Homo sapiens (man)

C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 09-Jul-2004

C:Accession: JC7506

R:McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hircock, M.; Pat

Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000

A:Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family me

A:Reference number: JC7506

A:Accession: JC7506

A:Molecule type: mRNA

A:Residues: 1-480 <MCK>

A:Cross-references: UNIPROT:Q9HB39; UNIPARC:UPI000003E88A; GB:AF282885

C:Comment: This protein, an intracellular membrane-bound enzyme, has biological and therat

C:Genetics:

A:Gene: hpa2a

A:Map position: 10q23-10q24

C:Keywords: heparin binding; membrane bound

Query Match	32.0%	Score	895.5	DB 2	Length	480			
Best Local Similarity	37.9%	Pred. No.	1.1e-61						
Matches	194	Conservative	70	Mismatches	141	Indels	107	Gaps	6
QY	33	LEPYTKRPLRSVSPSFLSITIDASLATDPRLFTFLGSPRLRALARGLSPAYLRFGGKTKTD	92						
Db	63	LDVSTKPNVTVNENFLSLQIDPSIHD-GWLDLFLSSKRLVTLARGLSPAFLRFGGKRTD	121						
QY	93	FLIFDPPDKPTSEERSYWKSVQNHDDICRSEPVSAAVLRKLQVWPPQELLLREQYQKEF	152						
Db	122	FLQFQNLNRPAS-----RGGPGP-----DYLKKNY	147						
QY	153	KNSTYSRSSVDMLYSFAKCSGLDLIFGLNALLRTPDLRWNSSNAQLLLDYCSKSGYNISW	212						
Db	148	E-----	148						
QY	213	ELGNPNPSFWKKAHLIDGLQGEDFVELHKLQR-SAFQNAKLYGPDIGQPRGKTVKLL	271						
Db	149	---DEPNRYTMHGRAVNSQLGKDYIQLKSLQIPRIYSRASLYGNIPRKNVIAL	205						
QY	272	RSFLKAGEVIDSLTWHHYLYNGRIATKEDFLSSDALDTFILSVOKILKVTKEITPGKKV	331						
Db	206	DGFVKYAGSTVDATWQHGYIDGRVVKYMDFLKTRLLDLSQIRKIQKVNTYTPGKKI	265						
QY	332	WLGETSSAYGGGAPLLSNTPAAGFMWLDKLGLSAQMGIEVVMRQVFFGAGNTHLVDENFE	391						
Db	266	WLEGVVTTSAGTNNLSDSYAAGFLWNLTLGLMANQGDVWIRHSFFDHGYNHLVDQNFN	325						
QY	392	PLPDYWLSSLPKKLVGRVLLSRVKGPD-----RSKLRVYLHCTNVYHPRVOEGDL	442						
Db	326	PLPDYWLSSLYKRLTIGPKVLAVHVGAGLQRPGRPRVIRDKLRIYAHCTNNHNNHYVRGSI	385						

[illegible]

## RESULT 2

T45608  
hypothetical protein F13G24.30 - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 09-Jul-2004  
C:Accession: T45608  
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba  
submitted to the Protein Sequence Database, December 1999  
A:Reference number: Z23009  
A:Accession: T45608  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-521 <BEV>  
A:A:Cross-references: UNIPROT:Q9SDA1; UNIPARC:UPI000000A497C; EMBL:AL133421  
A:Experimental source: cultivar Columbia; BAC clone F13G24  
C:Genetics:  
A:Map position: 5  
A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3  
A:Note: F13G24.30

	Query Match	14.0%;	Score 391;	DB 2;	Length 521;
	Best Local Similarity	28.4%;	Pred. No. 2.2e-22;		
	Matches 150; Conservative	73;	Mismatches 163;	Indels 122;	Gaps 24;
Qy	67	LGSPLRALARGSLPAVLRGCTKTDLFIDPDKEPTSEERSYWKSQVNHDICRSEPVA	126		
	:	:	:  :  :  :	:	
Dd	55	LTRPLLTKAIFKLIRIGSQLDQVIDVGNLKP-----CR-----	94		
Qy	127	AVLRKUQVEWPPFQELLRLREQYKFKNSS-----TVSRSSV-----DMLYSPAKCSGLDLIP	178		
	:	:	:  :  :  :	:	
Dd	95	-----PFQFM-----NSGLFGFSKGCLMKRWDELSFLTATGAVVF	132		

[illegible]

Qy	286	TWHHYLNGRATKEDFLSSDALD--TFILTSVQKILK----	VTKEITPGKVKWLGTSAY	344
Db	247	T-HIYNLG--SGNDPALVKKIMBSYLSQVSKTFKDVNQTIQHGHPWAS	PWVGSGGNY	303
Qy	341	GGGAPLLSNTFAAGFMWLDKLGSLAQMGIEVVMRQVFFGAGNYHLVDE--	NPEEPDLYWLS	399
Db	304	NSGGRHVSDFITDSFWYLDQLQMGARHNTKYYCQTLVGF-GFYGLLEKXGPNEDYSA	362	

Qy 400 LLPKLLVGPVRLSRVKGPPDRSKLRVYLHCTNVVHPYQEGDILLTVLNLH----- 455

Db 363 LLWHLRMGKGVLAQTQGP--PQLRVYAHCK-----GRAGVTLLNLINLSQSDFTVSV 414

Qy 452 -----VTXHLKVP-PPLFRKPYDTYL-----LKBSPGCLLI-SKSVQ 486

Db 415 SNGINVVNLNBSRKKSLDITLTKRPFSGWIGSKASGYLNREHYHTPE-SGLVSKINW 472

Qy 487 LGGQILKMWDEQTLPALTEKPLPAGSALSILPAFSYGGFFVIRNAKIAAC 534  
Db 473 LGGKILKPTATGDIPLSEPVLRSVNSPLNVLPLSMGFIVLPNFDASAC 520

### RESULT 3

T01953

hypothetical protein T2L5.6 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 09-Jul-2004  
C:Accession: T01953  
R:Geisel, C.; Smith, A.; Le, T.  
submitted to the EMBL Data Library, October 1998  
A:Description: The sequence of A. thaliana T2L5.  
A:Reference number: Z14470  
A:Accession: T01953  
A:Status: translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-190 <GEI>  
A:Cross-references: UNIPROT:O82604; UNIPARC:UPI00000A8F7D; EMBL:AF096371; NID:g3695386; E  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4  
A:Introns: 36/2; 69/3  
A:Note: T2L5.6  
C:Superfamily: Arabidopsis thaliana hypothetical protein T2L5.6

Query Match	5.8*	Score	161.5;	DB 2;	Length	190;				
Best Local Similarity	25.8*	Pred. No.	3.5e-05;							
Matches	50;	Conservative	35;	Mismatches	60;	Indels	49;	Gaps	8	
Qy	374	ROVFFGAGNTHLVD-ENFEP	LPD	YWL	SL	FKKL	GVPRVLLSRVKG	PD	SKLRVYLHCTNV	432
Db	12	RQSLIG-GNYGLNTWTF	NP	PDYYSALI	WQLMGR	KALFT	FSG	--TKK	IRSYTHCAR-	67
Qy	433	YHPRYQEGDI	TL	YVNLHNV	-----	-----	-----	-----	-----	475
Db	68	-----	QSGKITVLLMNL	NDNTT	TVVAK	VELNNS	FSLRHT	TKHKM	-----	115
Qy	476	GPDCGLL	-----	-----	-----	-----	-----	-----	-----	520
Db	116	GPFGVQIREEHVLTAK	DNLSHSQ	TMLNGALQ	NVSMG	DLPP	IEPIH	INST	ETIAPYS	175
Qy	521	YGFFVIRNAKIAAC								534
Db	176	IVFVHMENVVVPAC								189

## RESULT 4

mannan endo-1,4-beta-mannosidase (EC 3.2.1.78) - Caldocellum saccharolyticum  
A48954  
N:Alternate names: beta-mannanase  
C:Species: Caldocellum saccharolyticum  
C:Date: 19-Dec-1993 #sequence revision 18-Nov-1994 #text change 09-Jul-2004  
R:Accession: A48954; B43745  
R:Gibbs, M.D.; Saul, D.J.; Luthi, E.; Bergquist, P.L.  
Appl. Environ. Microbiol. 58, 3864-3867, 1992  
A:Title: The beta-mannanase from "Caldocellum saccharolyticum" is part of a multidomain  
A:Reference number: A48954; MUID:93119139; PMID:1476429  
A:Accession: A48954  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-1331 <GB>  
A:Cross-references: UNIPROT:P22533; UNIPARC:UPI000012EB88; GB:L01257; NID:g144290; PIDN:  
A>Note: sequence extracted from NCBI backbone (NCBIN:121576, NCBIPI:121577)  
R:Luthi, E.; Bhana Jasmal, N.; Grayling, R.A.; Love, D.R.; Bergquist, P.L.  
Appl. Environ. Microbiol. 57, 694-700, 1991  
A:Title: Cloning, sequence analysis, and expression in *Escherichia coli* of a gene coding  
A:Reference number: A43745; MUID:91247819; PMID:2039230

Query Match 4.2%; Score 117; DB 2; Length 1331;  
Best Local Similarity 19.1%; Pred. No. 1.9;

	Matches	108;	Conservative	72;	Mismatches	194;	Indels	192;	Gaps	24;
Qy	21	PAGTA-PTDDVVDLFEYTKRPLRSVSPSFLSGITIDASLATDPREFTLFGSLRLRAL---	A	76						
Db	759	FTVTATPTPTPTPIPTVTPLPITISPSVVEITINNAGRTQI-----SPIYGANQDI	812							
Qy	77	RGLSPAYLRFGGTK-----TDFLIPDDKEPTSEERSYKWSQNVDICRSEPVAAVL	129							
Db	813	EGVHSAARLGNGRLTYGNWENNFSNAGNDWYHSSDDYLCSMGISGEDAK---VPAAV	869							
Qy	130	RKLQVEMFQELLILLREYQEFKNSTYSRSSVDM-----	165							
Db	870	SKP-----HEYSLKNNAYSATQLQWGVSKDNYGTUSENETAPSNRAE	914							
Qy	166	YSFAKCGSLDILFGLNALLRPDLRWNSNAQLLLDYCSSKGYNIS-----WELGNE	217							
Db	915	VKFKKDAPLSL-----NPDLNDFVMDEFINYLINK-YGMASSPTGIKGIVLDNE	964							
Qy	218	PNSFWKKAHIIDG-----LOIGEDFVELHKLLORSAPONAKLYG-----	257							
Db	965	PD-LWASTHPRIHPNKVTKCELIKSVELAKVI-KTLDPSAEVFGYASYGFMGYSLQDA	1022							
Qy	258	PDIGQPRG-----KTVKLLRSFLKAGGEVIDSLTWHHVYVLNGRIATKESDFLSSD	306							
Db	1023	PDNRQVGGEHRWFISWLEQMCKKADSPFKLLDVLD-LHWYPARGGNIRVCFDGENDT	1081							
Qy	307	ALDTFIILSVQK-----ILKVTEITPGKKVLMGETSYAAGGAPLLSNT-----	350							
Db	1082	SKEVVIARMQAPRTLWDPTYKTSVKQGQITAGENSWINQWFSFY---LPIIENVKADIKY	1138							
Qy	351	-----FAAGFWMLDKLGUSAQWGIEVNM-----QVF---	377							
Db	1139	YPGTKLAISEPDYGGRNHISGIALADVLFGRKGVNFARWGDGSGSAAAAANYILNY	1198							
Qy	378	-----RGAGNYHLVDENFEPLDPLWLSLLPKLVGPRLVSRVKGPDRSKLVRLHCTN	431							
Db	1199	DGKSGKYGNTNVASNTSDVENMPVY-----ASINGQDDSELHILNRN	1242							
Qy	432	VYHPRYQEGDLTVLNLHNVTCHLK	457							
Db	1243	-YDQKIQ-----VKINITSPKYTK	1261							
RESULT	5									
T50113	3-dehydroquininate synthase (EC 4.2.3.4) - fission yeast (Schizosaccharomyces pombe)									
N;Contains:	3-dehydroquininate dehydratase; 3-dehydroquininate synthase (EC 4.6.1.3)									
C;Species:	Schizosaccharomyces pombe									
C;Date:	09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004									
C;Accession:	T50113									
R;Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.	submitted to the EMBL Data Library, February 2000									
A;Reference number:	Z25039									
A;Accession:	T50113									
A;Status:	preliminary; translated from GB/EMBL/DBJ									
A;Molecule type:	DNA									
A;Residues:	1-1573 <SEQ>									
A;Cross-references:	UNIPROT:Q9P7R0; UNIPARC:UIPO000125F22; EMBL:AL157734; PIDN:CAB75770									
A;Experimental source:	strain 972h(-); cosmid c1834									
C;Genetics:										
A;Gene:	SPDB:SPAC1834.02									
A;Map position:	1									
C;Superfamily:	pentafunctional Arom protein; 3-dehydroquininate dehydratase homology; 3-dehydroquininate kinase homology									
C;Keywords:	carbon-oxygen lyase; phosphorus-oxygen lyase									
F;407-835/Domain:	3-phosphoshikimate 1-carboxyvinyitransferase homology <PSK>									
F;1035-1279/Domain:	3-dehydroquininate dehydratase homology <DQD>									
Query Match	4.1%; Score 114.5; DB 2; Length 1573;									
Best Local Similarity	21.1%; Pred.No.3.8;									
Matches	120; Conservative 83; Mismatches 184; Indels 183; Gaps 32;									
Qy	18	QGAPAGTAPTDVVLDLEFYTKRPLRSVSPSFLSGITIDASLAT-----DP-RFLTFLGSPP	71							





Db 271 -----TKEG--PPDAKMFVLSVKILYVGM-----KKGNFQHSFLAGGA 310  
Qy 345 PLLSNTFAAGFMDKLGLSAQMGIEVVMROVPGAGNYHLVDENRPPDPYDYLWLSLLFKK 404  
Db 311 -----TUSAGRIYVD-----DGLKAVWPHSGHYLPTEENFOA-----FMSFLREN 351  
Qy 405 LVGPRVLLSRVKGPD 419  
Db 352 NVD---LANVKKNPD 363

RESULT 9  
RRNZAZ

Genome polyprotein - human respiratory syncytial virus (strain A2)  
N:Alternate names: polymerase L protein  
N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)  
C:Species: human respiratory syncytial virus  
A:Note: host Homo sapiens (man)  
C:Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text\_change 09-Jul-2004  
C:Accession: A40317; A28319; PS0048  
R:Stec, D.S.; Hill III, M.G.; Collins, P.L.  
Virology 183, 273-287, 1991  
A:Title: Sequence analysis of the polymerase L gene of human respiratory syncytial virus  
A:Reference number: A40317; MUID:91272488; PMID:2053282  
A:Accession: A40317  
A:Molecule type: mRNA  
A:Residues: 1-2165 <STE>  
A:Cross-references: UNIPROT:P28887; UNIPARC:UPI0000134AEA; GB:M75730; NID:G333955; PIDN:R:Collins, P.L.; Olmsted, R.A.; Spriggs, M.K.; Johnson, P.R.; Buckler-White, A.J. Proc. Natl. Acad. Sci. U.S.A. 84, 5134-5138, 1987  
A:Title: Gene overlap and site-specific attenuation of transcription of the viral polymerase  
A:Reference number: A28319; MUID:97260943; PMID:2440043  
A:Accession: A28319  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-81 <COL>  
A:Cross-references: UNIPARC:UPI0000134AE9; GB:M17245; NID:G333953; PIDN:AAA47417.1; PID:R:Johnson, P.R.; Collins, P.L. J. Gen. Virol. 69, 2901-2906, 1988  
A:Title: The A and B subgroups of human respiratory syncytial virus: comparison of inter  
A:Reference number: PS0048; MUID:89036169; PMID:3183631  
A:Accession: PS0048  
A:Molecule type: mRNA  
A:Residues: 1-18 <JOH>  
A:Cross-references: UNIPARC:UPI0000172722; GB:D00397; NID:G222551; PID:G2160375  
A:Experimental source: strain 18537  
A:Note: this strain belongs to subgroup B  
C:Genetics:  
A:Gene: L  
C:Superfamily: parainfluenza virus RNA-directed RNA polymerase  
C:Keywords: ATP; nucleotidyltransferase

Query Match 3.9%; Score 109.5; DB 1; Length 2165;  
Best Local Similarity 22.5%; Pred. No. 15;  
Matches 89; Conservative 48; Mismatches 119; Indels 139; Gaps 21;  
Qy 226 HILIDGLOIGEDFVELHKLQSAFONAKLYGPD-ICQPRGKTVKLSRSLKAGGEVIDS 284  
Db 1420 HLMKPPIFTGD--VDIHKLQ--VIOQHMFDPDKISLTQYVELFUSNKTLSGSHVNSN 1475  
Qy 285 LTHWH---YLLNGRIATKEDFLSSDALDFTILSVQKILKVTKEITPGKKVWLGETSSAY 340  
Db 1476 LILAHKISDYFHTYI-----LSTNLAGHWILLIQ-LMKDSKGIF--EKDW-GE----- 1520  
Qy 341 GGGAPLLSNTFAAGFMWLDKGLSAQMGIEVVMROVPGAGNYHLV-----D 387  
Db 1521 -----GY-----ITDHMFNL---KVFFNAYKTYLLCFHKGYGAKECD 1557  
Qy 388 ENFEPL-----PDWLSL-----LFKKLVGRVLLSRVKGPDRLKRLVYLHCTN 431  
Db 1558 MNTSDLLCVLELIDSSYWMKSMRVLEQKVIKYLSDASLHRVKGCHSFKL-WFLKRLN 1616  
Qy 432 V-----YHPRYQEGDIT-----LYVLNL--- 449

Db 1617 VAEFTVCPVVVNIYHPTMKAILTYIDLVRMGLINIDRIHIKNKHKNFDEFTSNLFYI 1676  
Qy 450 -----HNTYKHLKVP-PPLFRKPVDTYLLKPSGPDGLLSKSVOLNGQ-----ILK 493  
Db 1677 NYNFSNDNTHLLTKHIRANSLENNYNKLYHPTPETLENILANPIKSNDDKTKTNDYCI GK 1736  
Qy 494 MVDEQTLPALTEKPLPAGSALSPLAFS-----YGFF 524  
Db 1737 NVDSIMLPLLSNKKLIKSSAMIRTNYSKQDLYNLF 1771

## RESULT 10

G02529  
dynein heavy chain 1, cytosolic - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 09-Jul-2004  
C:Accession: G02529  
R:Vaisberg, E.A.; Griesom, P.M.; McIntosh, J.R.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: H01399  
A:Accession: G02529  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-897 <VAL>  
A:Cross-references: UNIPROT:Q14204; UNIPARC:UPI000016B1B7; EMBL:U53530; NID:G1314642; PI  
C:Superfamily: dynein heavy chain, cytosolic  
Query Match 3.8%; Score 105.5; DB 2; Length 897;  
Best Local Similarity 19.3%; Pred. No. 8.2;  
Matches 116; Conservative 85; Mismatches 206; Indels 195; Gaps 26;  
Qy 38 KRPLRSVSPSLSTIDASLATDPRFLTFLGSPRLRALARGLSPAYLRGGTKTDLIFD 97  
Db 219 EQPWVSQPRKLRQNLDAALLNQLKSF-----PARLRQYASVEYFVORLLKGYMKINMLVIE 273  
Qy 98 PDKEPTSEERSYKMSQVNHDIRSEPVSAALVKLOVMPFOELLL----- 143  
Db 274 LKSEALKDR--HWKO-----LMKRLHVNWVSVSELTGQIWDVDLQKNEAI 316  
Qy 144 -----LREQYQKEPKN--STYSRSVDMLYSFAKCSGLDLIFGLNALRTPD 189  
Db 317 VKDVLLVAQEMALEEFLKQIREVNTY--ELDLVNYQNKCR---LIRGWDLLFNKVK 370  
Qy 190 RWNSSNAQLLDYCSSKGYNI-----SWEIGNEPN-----SFW---KKAHILIDGL 234  
Db 371 HINSVSAMKLSPY-----YKVFEDALSWE--DKLNRIMALFDVWIDVQRRVWYLEGIFT 423  
Qy 235 GE-----DFVELHKLQSAFONAKLYGPDIGQPRGKTVKLSRFL 275  
Db 424 GSADIKHLLPVETORFOSISTEFLALMKVSKSLVMDVLNTQGVORSLERLADLLGKI 483  
Qy 276 KAGEVI--DSLTHWHYLANGRIATKEDFLSSDALDFTILSVQKILKVTKEITPGKKVWL 333  
Db 484 KALGEYLERERSSPRFYFVG-----DEDLLE-----IGNSKNVAKLQKH--KKMFA 530  
Qy 334 GETSSAYGGGAPLLSNTFAAGFMWLDKGLSAQMGIEVVMROVPGAGNYHLVDENFEPL 393  
Db 531 GVSSIIILNEDNSV-----LGISREGEVWPKTP-----VSITEHPK 568  
Qy 394 PDYWLSSLFKLVGRVLLSRVKGPDRLKRLVYLHCTN-----YHPRYQ----- 439  
Db 569 INEWLTLVEKEM--RVTLAKLAESVTEVEIFGKATSIDPNTYITWIDKYQAOQLVLSA 625  
Qy 440 -----GDLTLVYVLNHNVTYKHLKY-----PPPLFRKPEVD----- 468  
Db 626 QIAWSENVTALSSMGGGDAAPSDSVLSNVEVTLNVLADSVLMEQPPPLRRRKLHLEITE 685  
Qy 469 -----TYLLKPSGPDGLLSKSVOLNGQILKMVDQOTLPALTEKPLPAGSALSPLAFS 521  
Db 686 LVHORDVTRSLIKSIDN--AKSFEWLSQMRPYFDPKQTDVLQQLSIQWANA-----KFNY 739  
Qy 522 GF 523

Db 740 GF 741

RESULT 11

A38905

dynein heavy chain, cytosolic - rat

N;Contains: dynein ATPase (EC 3.6.4.2)

C;Species: Rattus norvegicus (Norway rat)

C;Date: 15-Apr-1994 #sequence\_revision 02-May-1994 #text\_change 09-Jul-2004

C;Accession: A38905; I58139

R;Zhang, Z.; Tanaka, Y.; Nonaka, S.; Aizawa, H.; Kawasaki, H.; Nakata, T.; Hirokawa, N. Proc. Natl. Acad. Sci. U.S.A. 90, 7928-7932, 1993

A;Title: The primary structure of rat brain (cytoplasmic) dynein heavy chain, a cytoplasmic dynein heavy chain

A;Reference number: A38905; MUID:93376715; PMID:7690137

A;Accession: A38905

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-4644 <ZHA>

A;Cross-references: UNIPROT:P38650; UNIPARC:UPI000013C4AA; GB:D13896; NID:g402527; PIDN:R1Mikami, A.; Paschal, B.M.; Mazumdar, M.; Vallee, R.B. Neuron 10, 787-796, 1993

A;Title: Molecular cloning of the retrograde transport motor cytoplasmic dynein MAP 1C.

A;Reference number: I58139; MUID:93264075; PMID:7684232

A;Accession: I58139

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1023, 'MP', 1026-1771, 'D', 1773-2097, 'A', 2099-2138, 'V', 2140-2174, 'A', 2176-2184 <RES>

A;Cross-references: UNIPARC:UPI000016867D; GB:I08505; NID:g294542; PIDN:AAA41103.1; PID:C;Superfamily: dynein heavy chain, cytosolic

C;Keywords: ATP; blocked amino end; heterotrimer; hydrolase; microtubule binding; nucleotide-binding motif A (P-loop)

F:1904-1911/Region: nucleotide-binding motif A (P-loop)

F:2222-2229/Region: nucleotide-binding motif A (P-loop)

F:2593-2600/Region: nucleotide-binding motif A (P-loop)

F:2935-2942/Region: nucleotide-binding motif A (P-loop)

F:1910/Binding site: ATP (lys) #status predicted

F:2228/Binding site: ATP (lys) #status predicted

F:2599/Binding site: ATP (lys) #status predicted

F:2941/Binding site: ATP (lys) #status predicted

Query Match 3.8%; Score 105.5; DB 1; Length 4644;

Best Local Similarity 19.3%; Pred. No. 96;

Matches 116; Conservative 84; Mismatches 207; Indels 195; Gaps 26;

Qy 38 KPLRSVSFSLITDASLATDPRFLTFLGSPRLALARGLSPAYLRFQGTDTDFLIPD 97

Db 1346 EQPWSVQPKRLQNLDGLNLQKNF-----PARLRQYASYEFVQRLKKGMYKINMLVIE 1400

Qy 98 PDKEPTSEERSYWKSVQVNHDCRSEPVSAAVLRKLOVEWPFQELLL----- 143

Db 1401 LKSEALKDR--HWKQ-----LMKRLHNVWVSELTGQIWDVDLQKNEAI 1443

Qy 144 -----LREYQKEFKN--STYRSVDMLYSFACSGLDLIFGLNALLRTPDL 189

Db 1444 VKDVLVAQGEALEBFLKQIREVMNTY---ELDLVNYQNKCR--LIRGWDLDLENKVK 1497

Qy 190 RWNSSNAQLLLDYCSSKGYNI-----SWELGNEPN-----SFW----KKAHLIDGLQL 234

Db 1498 HINSVSAMKLSPY-----YKVFEDALSWE--DKLARIMALFDVWIDVQRRWYLEGIFT 1550

Qy 235 GE-----DVEVHLKLLQRFSAQNAKLYGPDIGOPRCKTVKLLRSFL 275

Db 1551 GSADIKHLPLVETQRFQSTISTEFLAMKKVSKSPLVMDVNLNIQGVORSLERLADLIGKIQ 1610

Qy 276 KAGGEVI--DSLFWHHYLLNGRIATKEDFLSSDALDTFILSVOKILKVTKEITPGKKVWL 333

Db 1611 KALGEVLERSERSFPFYVG-----DEDLLE-----IIGNSKNVAKLQKHF---KKMFA 1657

Qy 334 GETSSAYGGAPLLSNTFFAAGFMWLDKLGLSAQMGIEVVMRQVFFGAGNTHLVDENPFL 393

Db 1658 GVSSIIILNDSVV-----LGTSSRGESEVMFKTP-----VSITEHPK 1695

Qy 394 PDYWLSSLFKKLVGPRVLLSRVKGDPDRSKLRVYLHCT-----NVYHPRVOEGDLTYVLN- 448

Db 1696 INEWLTLVEKEM---RVTAKLLAESVTEVEIFGKATSIDENTYITWIDKYOQLWLSA 1752

Qy 449 -----LHNVYKHLKV-----PPPLFRKPV----- 468

Db 1753 QIAWSENVENALSNVGGGNGVPLQSVLSNVSEVTLNVLADSVLMQPPPLRRRKLHLE 1812

Qy 469 -----TYLLKPSGPDGLLSKSVQLNGQILKVMDEOTLPALTEKPLPAGSALSLPATSY 521

Db 1813 LVHQRDVTRSLIKSKIDN--AKSFEMLSQRFYDPKQTDVLQQLSIQMANA-----KFN 1866

Qy 522 GF 523

Db 1867 GF 1868

RESULT 12

S36231

beta-fructofuranosidase (EC 3.2.1.26) - potato (fragment)

N;Alternate names: invertase

C;Species: Solanum tuberosum (potato)

C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004

C;Accession: S36231; S34145

R;Hedley, P.E.; Machray, G.C.; Davies, H.V.; Burch, L.; Waugh, R. Plant Mol. Biol. 22, 917-922, 1993

A;Title: cDNA cloning and expression of a potato (Solanum tuberosum) invertase.

A;Reference number: S36231; MUID:93363925; PMID:8358038

A;Accession: S36231

A;Molecule type: mRNA

A;Residues: 1-587 <HED>

A;Cross-references: UNIPROT:Q43172; UNIPARC:UPI0000175B36; EMBL:Z21486

R;Hedley, P.E.; Machray, G.C.; Davies, H.V.; Burch, L.; Waugh, R. submitted to the EMBL Data Library, January 1993

A;Reference number: S34145

A;Molecule type: mRNA

A;Residues: 1-576, 'KM', 579-587 <HE2>

A;Cross-references: UNIPARC:UPI00000AD071; EMBL:Z21486; NID:g313128; PIDN:CAA79676.1; PIIC;Superfamily: beta-fructofuranosidase

C;Keywords: glycosidase; hydrolase

Query Match 3.7%; Score 104.5; DB 2; Length 587;

Best Local Similarity 26.2%; Pred. No. 5.2;

Matches 43; Conservative 22; Mismatches 54; Indels 45; Gaps 8;

Qy 77 RGLSPAYLRFEGTGTDFLIFFDPOKEPTSEERSYW-----KSQVNHDCRSEPVSAAVLRK 131

Db 306 KGLRDYGYNTYASKS---FYDPSK---NRRIMGMWANESDTVNDVKKGWAGIQTIPRK 358

Qy 132 L-----QVEWPFQELLLREYQK-----EFKNSTYSRSSVDMLYSPAKC 171

Db 359 LMLDPSKGQVLQWPFVEELETIREQVQLSNRKLKKGDKIEVKGITPAQADVETFSF--- 415

Qy 172 SGLDLIFGLNALLRTPDLRWNSSNAQLLLDYCSSKGYNISWELG 215

Db 416 SSLD-----KAEPFDPNDNLVYQ---DVCAIKGSTVQGDIG 449

RESULT 13

G72395

alpha-L-arabinofuranosidase - Thermotoga maritima (strain MSBs)

C;Species: Thermotoga maritima

C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004

C;Accession: G72395

R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M. Nature 399, 323-329, 1999

A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing

A;Reference number: A72200; MUID:99287316; PMID:10360571

A;Accession: G72395

A;Status: preliminary





GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model  
Run on: June 5, 2006, 12:01:06 ; Search time 106.661 Seconds  
(without alignments)  
2293.354 Million cell updates/sec

Title: US-10-645-659A-2  
Perfect score: 2797  
Sequence: 1 MLRLLLWLWGLGALAQA.....LPAPSYGFFVIRNAKIAACI 535

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_8.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*
- 10: Geneseqp2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2797	100.0	535	3 AAB08851	Aab08851 A murine
2	2797	100.0	535	5 ABB07811	Abb07811 Mouse hep
3	2797	100.0	535	7 ADG88834	Adg88834 Mouse hpa
4	2797	100.0	535	8 ADL16413	Adl16413 Mouse hep
5	2797	100.0	535	8 ADM48750	Adm48750 Mouse hpa
6	2797	100.0	535	8 ADR88208	Adr88208 Mouse hep
7	2797	100.0	535	8 ADT78175	Adt78175 Mouse hep
8	2797	100.0	535	9 AEA42424	Aea42424 Mouse hep
9	2793	99.9	535	9 ADY27033	Ady27033 Murine he
10	2590.5	92.6	536	5 ABB07812	Abb07812 Rat hep
11	2590.5	92.6	536	8 ADR88209	Adr88209 Rat hep
12	2590.5	92.6	536	8 ADT78176	Adt78176 Rat hep
13	2590.5	92.6	536	9 ADY27035	Ady27035 Rat hep
14	2590.5	92.6	536	9 AEA42425	Aea42425 Rat hep
15	2184	78.1	545	5 ADY27034	Ady27034 Bovine he
16	2149	76.8	543	2 AAY17082	Aay17082 Human hep
17	2149	76.8	543	4 AAB86206	Aab86206 Human hep
18	2149	76.8	543	7 ADD18950	Add18950 Human dis
19	2149	76.8	543	8 ADK52086	Adk52086 Human ato
20	2149	76.8	543	8 ADM48759	Adm48759 Human hpa
21	2149	76.8	543	8 ADN05074	Adn05074 Antipsori
22	2149	76.8	543	8 ADN04902	Adn04902 Antipsori
23	2149	76.8	543	8 ADQ80372	Adq80372 Heparanas

24	2149	76.8	543	8 ADR88210	Adr88210 Human pre
25	2149	76.8	543	8 ADP25079	Adp25079 PRO polyv
26	2149	76.8	543	8 ADT78177	Adt78177 Human hep
27	2149	76.8	543	9 ADY27036	Ady27036 Human hep
28	2149	76.8	543	9 AEA42426	Aea42426 Human hep
29	2149	76.8	588	2 AAY30124	Aay30124 A human p
30	2147.5	76.8	545	6 ABP56822	Abp56822 Human hep
31	2147.5	76.8	545	7 ADE16012	Ade16012 G-coupled
32	2147.5	76.8	545	8 ADL93951	Adl93951 Human G-c
33	2146	76.7	543	2 AAY02345	Aay02345 A human h
34	2146	76.7	543	3 AAY57590	Aay57590 Human hep
35	2146	76.7	543	3 AAB08849	Aab08849 Amino aci
36	2146	76.7	543	3 AAY52990	Aay52990 Human hep
37	2146	76.7	543	4 AAY97635	Aay97635 Human hep
38	2146	76.7	543	5 ABB07813	Abb07813 Human hep
39	2146	76.7	543	7 ADG88800	Adg88800 Human hpa
40	2146	76.7	543	8 ADL16379	Adl16379 Human hep
41	2146	76.7	543	8 ADM48716	Adm48716 Human hpa
42	2146	76.7	543	9 AEA42466	Aea42466 Human hep
43	2146	76.7	543	10 AEE96848	Aee96848 Human hep
44	2146	76.7	592	2 AAY02346	Aay02346 A human h
45	2146	76.7	592	3 AAB08850	Aab08850 Amino aci

ALIGNMENTS

RESULT 1

AAB08851	ID	AAB08851	standard; protein; 535 AA.
AC	AAB08851;		
XX			
DT	15-JAN-2001	(first entry)	
XX			
DE	A murine heparanase polypeptide.		
XX			
KW	Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;		
KW	heparin-binding growth factor; cytokine; neurodegenerative plaque;		
KW	wound healing; infection; burn; angiogenesis; restenosis;		
KW	atherosclerosis; inflammation; neurodegenerative disease;		
KW	Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease.		
XX			
OS	Mus sp.		
XX			
PN	WO200052178-A1.		
XX			
PD	08-SEP-2000.		
XX			
PF	14-FEB-2000; 2000WO-US003542.		
XX			
PR	01-MAR-1999; 99US-00258892.		
XX			
PA	(INSI-) INSIGHT STRATEGY & MARKETING LTD..		
PA	(HADA-) HADASIT MEDICAL RES SERVICES & DEV.		
PA	(FRIE/) FRIEDMAN M M.		
XX			
PI	Pecker I, Vlodavsky I, Feinstein E;		
XX			
DR	WPI; 2000-579289/54.		
DR	N-ESDB; AAA75081.		
XX			
PT	New polynucleotides encoding a polypeptide having heparanase activity,		
PT	useful in wound healing and in gene therapy, particularly in treating		
PT	tumor, inflammation, autoimmunity, neurodegenerative diseases.		
XX			
XX	Claim 22; Page 144-145; 152pp; English.		
XX			
CC	The present sequence represents murine protein with heparanase catalytic		
CC	activity. The heparanase (hpa) polynucleotide is useful in gene therapy,		
CC	particularly in treating tumour, inflammation or autoimmunity.		
CC	Particularly, the polynucleotide is useful in modulating the		
CC	bioavailability of heparin-binding growth factors, cellular responses to		

CC heparin-binding growth factors (e.g. bFGF) and cytokines (e.g. interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular susceptibility to certain viral and some bacterial and protozoa infections, or disintegration of neurodegenerative plaques. The polynucleotide is also useful in wound healing (e.g. thermal, chemical or radiation burns), and in the treatment of angiogenesis, restenosis, atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-Sträussler Syndrome or Creutzfeldt-Jakob disease), and some viral, bacterial or protozoa infections

XX Sequence 535 AA;

Query Match 100.0%; Score 2797; DB 3; Length 535;  
Best Local Similarity 100.0%; Pred. No. 2.9e-262;  
Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRLLLWLWGPILGALAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPSLTITDASLATD 60  
DB 1 MLRLLLWLWGPILGALAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPSLTITDASLATD 60

QY 61 PRFLTFLGSPRLRALARGLSPAYLRFGGTKTDFLIFDPDKPTSEERSYWKSONVHDI 120  
DB 61 PRFLTFLGSPRLRALARGLSPAYLRFGGTKTDFLIFDPDKPTSEERSYWKSONVHDI 120

QY 121 SEPVSAAVLRKQLQVEWPFQELLRLREQYQKEFKNSTYSRSSVDMLYSFACSGLDLIFGL 180  
DB 121 SEPVSAAVLRKQLQVEWPFQELLRLREQYQKEFKNSTYSRSSVDMLYSFACSGLDLIFGL 180

QY 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHLIDGLQGEDFVE 240  
DB 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHLIDGLQGEDFVE 240

QY 241 LHKLLQRSFONAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300  
DB 241 LHKLLQRSFONAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300

QY 301 DFLSSDALDTFILSVQKILKVTKEITPGKKVWLGETSSAYCGGAPLLSNTFAAGFMWLDK 360  
DB 301 DFLSSDALDTFILSVQKILKVTKEITPGKKVWLGETSSAYCGGAPLLSNTFAAGFMWLDK 360

QY 361 LGLSAQMGIEVVMRQVFFGAGNTHLVNDENPEPLPDYWLSSLFVKLVGPRVLLSRVKGPD 420  
DB 361 LGLSAQMGIEVVMRQVFFGAGNTHLVNDENPEPLPDYWLSSLFVKLVGPRVLLSRVKGPD 420

QY 421 SKLRVYLHCTNVVHPRYQEGDLTYVLNLHNVTKHLKVPDPVDTYLLKPSGPDGL 480  
DB 421 SKLRVYLHCTNVVHPRYQEGDLTYVLNLHNVTKHLKVPDPVDTYLLKPSGPDGL 480

QY 481 LSKSVQLNGQILKMVDQTLTPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAACI 535  
DB 481 LSKSVQLNGQILKMVDQTLTPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAACI 535

RESULT 2

ID ABB07811 standard; protein; 535 AA.

XX ABB07811;

AC ABB07811;

DT 03-JUL-2002 (first entry)

DE Mouse heparanase sequence.

XX Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;

KW anti-protozoan; neuroprotective; heparin; mouse.

XX Mus musculus.

Key Location/Qualifiers

Peptide 1..17

FT /note= "putative signal peptide"

FT 18..535

FT /note= "mature protein"

XX US2002034810-A1.  
XX 21-MAR-2002.  
XX 16-AUG-2001; 2001US-00930218.  
XX 20-SEP-2000; 2000US-00666390.  
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;  
XX WPI; 2002-338926/37.  
XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful  
XX to treat various heparin-related disorders and the signal peptide is  
XX useful in production of membrane-targeted or secreted recombinant  
XX proteins.  
XX Disclosure; Fig 1a; 39pp; English.  
XX The invention relates to an isolated avian and reptile nucleic acid,  
XX encoding a polypeptide with heparanase catalytic activity. The signal  
XX peptide of the nucleic acid can be used to express membrane-associated or  
XX secreted proteins in heterologous expression systems. The encoded  
XX polypeptides can be used to prevent tumour angiogenesis, metastasis and  
XX invasion, and to intervene with pathologies associated with impaired  
XX heparin-binding growth factors, cellular responses to heparin-binding  
XX growth factors and cytokines, cell interaction with plasma lipoproteins,  
XX cellular susceptibility to viral, protozoa and bacterial infections or  
XX disintegration of neurodegenerative plaques. The present sequence  
XX represents a mouse heparanase protein sequence used in similarity studies  
XX Sequence 535 AA;

Query Match 100.0%; Score 2797; DB 5; Length 535;  
Best Local Similarity 100.0%; Pred. No. 2.9e-262;  
Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRLLLWLWGPILGALAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPSLTITDASLATD 60  
DB 1 MLRLLLWLWGPILGALAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPSLTITDASLATD 60

QY 61 PRFLTFLGSPRLRALARGLSPAYLRFGGTKTDFLIFDPDKPTSEERSYWKSONVHDI 120  
DB 61 PRFLTFLGSPRLRALARGLSPAYLRFGGTKTDFLIFDPDKPTSEERSYWKSONVHDI 120

QY 121 SEPVSAAVLRKQLQVEWPFQELLRLREQYQKEFKNSTYSRSSVDMLYSFACSGLDLIFGL 180  
DB 121 SEPVSAAVLRKQLQVEWPFQELLRLREQYQKEFKNSTYSRSSVDMLYSFACSGLDLIFGL 180

QY 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHLIDGLQGEDFVE 240  
DB 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHLIDGLQGEDFVE 240

QY 241 LHKLLQRSFONAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300  
DB 241 LHKLLQRSFONAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300

QY 301 DFLSSDALDTFILSVQKILKVTKEITPGKKVWLGETSSAYCGGAPLLSNTFAAGFMWLDK 360  
DB 301 DFLSSDALDTFILSVQKILKVTKEITPGKKVWLGETSSAYCGGAPLLSNTFAAGFMWLDK 360

QY 361 LGLSAQMGIEVVMRQVFFGAGNTHLVNDENPEPLPDYWLSSLFVKLVGPRVLLSRVKGPD 420  
DB 361 LGLSAQMGIEVVMRQVFFGAGNTHLVNDENPEPLPDYWLSSLFVKLVGPRVLLSRVKGPD 420

QY 421 SKLRVYLHCTNVVHPRYQEGDLTYVLNLHNVTKHLKVPDPVDTYLLKPSGPDGL 480  
DB 421 SKLRVYLHCTNVVHPRYQEGDLTYVLNLHNVTKHLKVPDPVDTYLLKPSGPDGL 480

QY 481 LSKSVQLNGQILKMVDQTLTPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAACI 535



Db 481 LSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSPLAFSGFFVIRNAKIAACI 535  
|||||  
RESULT 3  
ADG88834  
ID ADG88834 standard; protein; 535 AA.  
XX AC ADG88834;  
XX DT 11-MAR-2004 (first entry)  
XX DE Mouse hpa protein.  
XX KW Wound healing; heparanase; ulcer; burn; laceration; surgical incision;  
XX KW necrosis; pressure wound; diabetic ulcer; angiogenesis; mouse; therapy.  
XX OS Mus musculus.  
XX PN US2003161823-A1.  
XX PD 28-AUG-2003.  
XX PF 14-JAN-2003; 2003US-00341582.  
XX PR 31-AUG-1998; 98WO-US017954.  
XX PR 01-MAR-1999; 99US-00258892.  
XX PR 06-FEB-2001; 2001US-00776874.  
XX PR 05-SEP-2001; 2001WO-IL000830.  
XX PR 19-NOV-2001; 2001US-00988113.  
XX (ILAN/) ILAN N.  
XX (VLOD/) VLODAVSKY I.  
XX (YACO/) YACOBY-ZEEVI O.  
XX (PECK/) PECKER I.  
XX (FEIN/) FEINSTEIN E.  
XX Ilan N, Vlodavsky I, Yacoby-Zeevi O, Pecker I, Feinstein E;  
XX WPI; 2003-897910/82.  
XX N-PSDB; ADG88833, ADG88835.  
XX Composition for treating a wound comprising recombinant heparanase is  
XX useful to induce or accelerate wound healing and induce or accelerate  
XX angiogenesis.  
XX Claim 2; SEQ ID NO 44; 143pp; English.  
XX The present invention relates to methods and compositions for inducing  
XX and/or accelerating wound healing via the catalytic activity of  
XX heparanase. The invention is used to induce or accelerate a healing  
XX process, particularly of an ulcer, burn, laceration, surgical incision,  
XX necrosis, pressure wound, diabetic ulcer and to induce or accelerate  
XX angiogenesis. The present sequence is mouse hpa protein.  
XX SQ Sequence 535 AA;  
Query Match 100.0%; Score 2797; DB 7; Length 535;  
Best Local Similarity 100.0%; Pred. No. 2.9e-262;  
Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MLRLLLMLWGPICGALAQGAPAGTDDVVDFEYTKRPLRSVSPSFLSITIDASLATD 60  
Db 1 MLRLLLMLWGPICGALAQGAPAGTDDVVDFEYTKRPLRSVSPSFLSITIDASLATD 60  
Qy 61 PRFLTFLGSPRLRALARGLPAYLRFGGTKTDFLIFDPDKPTSEERSYMKSQVNHDIR 120  
Db 61 PRFLTFLGSPRLRALARGLPAYLRFGGTKTDFLIFDPDKPTSEERSYMKSQVNHDIR 120  
Qy 121 SEPVSAAVLRKLQVWEPFQELLRLLEQYQKEFKNSTYSRSSVDMLYSPAKCSGLDLIFGL 180  
Db 121 SEPVSAAVLRKLQVWEPFQELLRLLEQYQKEFKNSTYSRSSVDMLYSPAKCSGLDLIFGL 180

Qy 181 NALLRTTDLRWSSNAQLLLDYCSSKGYNISWELGNPNFSFWKKAHILIDGLQGEDFVE 240  
Db 181 NALLRTTDLRWSSNAQLLLDYCSSKGYNISWELGNPNFSFWKKAHILIDGLQGEDFVE 240  
Qy 241 LHKLLQSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300  
Db 241 LHKLLQSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300  
Qy 301 DFLSSDALDTFILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTPAAGFMWLDK 360  
Db 301 DFLSSDALDTFILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTPAAGFMWLDK 360  
Qy 361 LGLSAQMGIEVVMQVRFAGNYHLVDENFPLPDYWLSSLPKLVGPRVLLSRVKGPD 420  
Db 361 LGLSAQMGIEVVMQVRFAGNYHLVDENFPLPDYWLSSLPKLVGPRVLLSRVKGPD 420  
Qy 421 SKLRVYLHCTNVYHPRYQEGDLTLYVLNLHNVTKHLKVPPLFRKPVDVTLKPSGPDGL 480  
Db 421 SKLRVYLHCTNVYHPRYQEGDLTLYVLNLHNVTKHLKVPPLFRKPVDVTLKPSGPDGL 480  
Qy 481 LSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSPLAFSGFFVIRNAKIAACI 535  
Db 481 LSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSPLAFSGFFVIRNAKIAACI 535  
RESULT 4  
ADL16413  
ID ADL16413 standard; protein; 535 AA.  
XX AC ADL16413;  
XX DT 06-MAY-2004 (first entry)  
XX DE Mouse heparanase protein.  
XX KW Mouse; heparanase; enzyme; heparanase-dependent cancer; cancer;  
XX KW autoimmune reaction; inflammation.  
XX OS Mus musculus.  
XX PN US2003236215-A1.  
XX PD 25-DEC-2003.  
XX PF 09-JUN-2003; 2003US-00456573.  
XX PR 31-AUG-1998; 98WO-US017954.  
XX PR 01-MAR-1999; 99US-00258892.  
XX PR 08-NOV-1999; 99US-00435739.  
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
XX Pecker I, Vlodavsky I, Feinstein E;  
XX WPI; 2004-070610/07.  
XX New antisense oligonucleotide hybridizable with a polynucleotide encoding  
XX a polypeptide with heparanase activity, useful for treating diseases such  
XX as cancer and autoimmune disorders.  
XX Claim 3; SEQ ID NO 44; 108pp; English.  
XX The invention relates to an antisense oligonucleotide (ASO) comprising a  
XX polynucleotide or a polynucleotide analogue of at least 10 bases being  
XX hybridizable in vivo, under physiological conditions, with a portion of  
XX a polynucleotide strand encoding a polypeptide having heparanase  
XX catalytic activity. Also included are a method of in vivo downregulating  
XX heparanase activity (comprising administering the ASO in vivo), a method  
XX of treating a subject suffering from a pathological condition  
XX (characterised by heparanase activity, comprising administering ASO to  
XX the subject), a pharmaceutical composition comprising the ASO and a  
XX carrier, an antisense nucleic acid construct (comprising a promoter

CC sequence and a polynucleotide sequence directing the synthesis of an  
 CC antisense RNA sequence of at least 10 bases being hybridisable in vivo,  
 CC under physiological conditions, with a polynucleotide strand encoding a  
 CC polypeptide having heparanase catalytic activity), a method of in vivo  
 CC downregulating heparanase activity (comprising administering in vivo the  
 CC antisense nucleic acid construct), a pharmaceutical composition  
 CC comprising the antisense nucleic acid construct and a carrier, and an  
 CC antisense oligonucleotide comprising a polynucleotide or a polynucleotide  
 CC analogue of at least 10 bases being hybridisable in vivo, under  
 CC physiological conditions, with a portion of a polynucleotide strand being  
 CC characterised by forming at least a portion of an untranslated region  
 CC (UTR) for a polynucleotide strand encoding a polypeptide having  
 CC heparanase catalytic activity. The methods and compositions of the  
 CC present invention are useful for the prevention and/or treatment of  
 CC diseases or conditions associated with aberrant heparanase activity, such  
 CC as heparanase-dependent cancer, cancer, autoimmune reaction and  
 CC inflammation. The gene for human heparanase is located on chromosome 4.  
 CC The present sequence is the mouse heparanase protein.

XX Sequence 535 AA;

Query Match 100.0%; Score 2797; DB 8; Length 535;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-262;  
 Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRLLLLWLGPLGALAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPSFLSITIDASLATD 60  
 DB 1 MLRLLLLWLGPLGALAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPSFLSITIDASLATD 60

QY 61 PRFLTFLGSPRLRALARGSPAYLRFGGKTDFLIFDPDKPTSEERSYKWSQVNHICR 120  
 DB 61 PRFLTFLGSPRLRALARGSPAYLRFGGKTDFLIFDPDKPTSEERSYKWSQVNHICR 120

QY 121 SEPVSAAVLRLKQVEWPFQELLRLREYQKEFNKSTYSRSSVDMLYSFACSGLDLIFGL 180  
 DB 121 SEPVSAAVLRLKQVEWPFQELLRLREYQKEFNKSTYSRSSVDMLYSFACSGLDLIFGL 180

QY 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNPNFSFWKKAHILIDGLQGEDFVE 240  
 DB 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNPNFSFWKKAHILIDGLQGEDFVE 240

QY 241 LHKLLORSFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYHLYNGRIATKE 300  
 DB 241 LHKLLORSFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYHLYNGRIATKE 300

QY 301 DFLSSDALDTFILSVQKILKVTKEITPGKKVWLGETSAYGGCAPILLSNTFAAGFWLDDK 360  
 DB 301 DFLSSDALDTFILSVQKILKVTKEITPGKKVWLGETSAYGGCAPILLSNTFAAGFWLDDK 360

QY 361 LGLSAQMGLIEVVRQVFFGAGNHYLDENPEPLPDYWLSSLFKLVGPRVLLSRVKGPD 420  
 DB 361 LGLSAQMGLIEVVRQVFFGAGNHYLDENPEPLPDYWLSSLFKLVGPRVLLSRVKGPD 420

QY 421 SKLRYVLHCTNVYHPRYQEGDLTYVNLHNVTKHLKVPPLPRKPVDTYLLKPSGPDGL 480  
 DB 421 SKLRYVLHCTNVYHPRYQEGDLTYVNLHNVTKHLKVPPLPRKPVDTYLLKPSGPDGL 480

QY 481 LSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAICI 535  
 DB 481 LSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSIPAFSYGFFVIRNAKIAICI 535

RESULT 5

ADM48750

ID ADM48750 standard; protein; 535 AA.

XX

AC ADM48750;

XX

DT 03-JUN-2004 (first entry)

XX

DE Mouse hpa protein.

XX

KW Transgenic animal; heparanase; cancer; viral infection; restenosis;

XX

KW neurodegenerative disease; atherosclerosis; pulmonary disorder; hpa;  
 KW mouse.

XX Mus musculus.

XX US2003217375-A1.

XX 20-NOV-2003.

XX 24-FEB-2003; 2003US-00371218.

XX 31-AUG-1998; 98WO-US017954.

XX 01-MAR-1999; 99US-00258892.

XX 08-FEB-2001; 2001US-00776874.

XX 19-NOV-2001; 2001US-00988113.

XX (ZCHA/) ZCHARIA E.

XX (VLOD/) VLODAVSKY I.

XX (METZ/) METZGER S.

XX (PECK/) PECKER I.

XX (ILAN/) ILAN N.

XX (CHAJ/) CHAJEK-SHAUL T.

XX (GOLD/) GOLDSHMIDT O.

XX Zcharia E, Vlodaevsky I, Metzger S, Pecker I, Ilan N;

XX Chajek-Shaul T, Goldshmidt O;

XX WPI; 2004-021918/02.

XX N-PSDB; ADM48749, ADM48751.

XX

XX New transgenic non-human animal expressing heparinase, useful as models

XX for human disease, such as cancers, viral infection, neurodegenerative

XX diseases, restenosis, atherosclerosis and pulmonary disorders.

XX

XX Example 12; SEQ ID NO 44; 106pp; English.

XX

XX The present invention relates to a transgenic non-human animal whose

XX genome comprises an exogenous polynucleotide sequence, including a

XX promoter active in tissues of the non-human, a region encoding a human

XX heparanase, where the promoter and the region encoding human heparanase

XX are operably linked in the exogenous polynucleotide such that human

XX heparanase is expressed in at least a portion of the cells of the non-

XX human animal. The methods and compositions of the present invention are

XX useful for the production of transgenic animals expressing heparanase, to

XX be used as models for human diseases such as cancers, viral infection,

XX restenosis, neurodegenerative diseases, atherosclerosis and pulmonary

XX disorders. The present sequence is mouse hpa protein used in the

XX exemplification of the invention.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Db 241 LHKLLQSAFQNAKLYGPDIGPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300  
 QY 301 DFLSSDALDTFILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDK 360  
 Db 301 DFLSSDALDTFILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDK 360  
 QY 361 LGLSAQMGIEVVMQVFFGAGNYHLVDENPEPLDYWLSSLFVKLVGPRVLLSRVKGPD 420  
 Db 361 LGLSAQMGIEVVMQVFFGAGNYHLVDENPEPLDYWLSSLFVKLVGPRVLLSRVKGPD 420  
 QY 421 SKLRVYLHCTNVYHPRYQEGDLTYLVNLHNVTKHLKVPPLFRKPVDTYLLKPSGPDGL 480  
 Db 421 SKLRVYLHCTNVYHPRYQEGDLTYLVNLHNVTKHLKVPPLFRKPVDTYLLKPSGPDGL 480  
 QY 481 LSKSVQLNGQILKXWVDSQTLPALTEKPLPAGSALSPLAFSGVFFVIRNAKIAACI 535  
 Db 481 LSKSVQLNGQILKXWVDSQTLPALTEKPLPAGSALSPLAFSGVFFVIRNAKIAACI 535

## RESULT 6

ADR88208  
 ID ADR88208 standard; protein; 535 AA.

AC ADR88208;

DT 18-NOV-2004 (first entry)

DE Mouse heparanase.

Targeted drug delivery; inflammatory disorder; wound; scar; vasculopathy;  
 autoimmune disorder; cancer; angiogenesis; metastatic disease;  
 atherosclerosis; restenosis; aneurysm; solid cancer; non-solid cancer;  
 haematopoietic malignancy; lymphocytic leukaemia; myelogenous leukaemia;  
 Hodgkin's disease; multiple myeloma; haemangiosarcoma; Kaposi's sarcoma;  
 mouse; heparanase; enzyme.

OS Mus musculus.

Key Location/Qualifiers  
 FT Peptide 1..17  
 FT Protein 18..535  
 FT /label= Signal\_peptide  
 FT /label= Mature\_heparanase

XX US2004170631-A1.

XX 02-SEP-2004.

XX 28-NOV-2003; 2003US-00722502.

XX 02-SEP-1997; 97US-00922170.

PR 01-MAY-1998; 98US-00071739.

PR 04-NOV-1998; 98US-00186200.

PR 19-FEB-2003; 2003US-00368044.

PR 22-AUG-2003; 2003US-00645659.

XX (YACO/) YACOBY-ZEEVI O.

PA (PERE/) PERETZ T.

PA (MIRO/) MIRON D.

PA (SHLO/) SHLOMI Y.

PA (PECK/) PECKER I.

PA (AYAL/) AYAL-HERSHKOVITZ M.

PA (FEIN/) FEINSTEIN E.

PA (VGEL/) VAN GELDER J M.

PA (VLOD/) VLODAVSKY I.

PA (FRIE/) FRIEDMANN Y.

XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;

PI Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;

PI Friedmann Y;

XX WPI; 2004-625084/60.

XX

Targeted drug delivery to a heparanase-expressing tissue of a patient, useful for treating heparanase-associated conditions such as inflammation or cancer, comprises administering a drug and an anti-heparanase antibody complex.

Claim 2; SEQ ID NO 2; 58pp; English.

The invention relates to a method of targeted drug delivery to a tissue of a patient, the tissue expressing heparanase. The method comprises providing a complex of a drug directly or indirectly linked to an anti-heparanase antibody, and administering the complex to the patient. In the targeted drug delivery, the antibody comprises an antibody or its portion capable of specifically binding to at least one epitope of a heparanase protein. The composition and methods of the invention are useful for diagnosing, preventing or treating conditions associated with heparanase catalytic activity (e.g. an inflammatory disorder, wound, scar, vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell proliferation, invasion of circulating tumour cells and metastatic disease), for purifying heparanase, or for developing drugs for those heparanase-associated conditions. The vasculopathy is atherosclerosis, restenosis or aneurysm. The cancerous condition is a solid cancer or a non-solid cancer. The non-solid cancer is a haematopoietic malignancy selected from acute lymphocytic leukaemia (ALL), acute myelogenous leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia, Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and multiple myeloma. The solid cancer is selected from tumours in lip and oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands, thyroid gland, oesophagus, stomach, small intestine, colon, colorectum, anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva, vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva, malignant melanoma of the conjunctiva, malignant melanoma of the uvea, retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit, brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's sarcoma. The present sequence is mouse heparanase.

Sequence 535 AA;

Query Match 100.0%; Score 2797; DB 8; Length 535;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-262;  
 Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRLLLWLGPIGALAQAGAPAGTDDVDLEFVTKPLRSVSPSFLSIITDASLATD 60  
 Db 1 MLRLLLWLGPIGALAQAGAPAGTDDVDLEFVTKPLRSVSPSFLSIITDASLATD 60  
 QY 61 PRFLTFLGSPRLRALRGLSPAYLRFGGTKTDFLI FDPDKPTSEERSYWKSVQVNHDIR 120  
 Db 61 PRFLTFLGSPRLRALRGLSPAYLRFGGTKTDFLI FDPDKPTSEERSYWKSVQVNHDIR 120  
 QY 121 SEPVSAAVLRKLQVEWPFQELLRLREQYQKFKNSTYSRSSVDMLYSFAKSGLDLIFGL 180  
 Db 121 SEPVSAAVLRKLQVEWPFQELLRLREQYQKFKNSTYSRSSVDMLYSFAKSGLDLIFGL 180  
 QY 181 NALLRTPDLRWNSSNAQLLLDYCSSKGNYSWELGNPNPSFWKKAHILIDGLQGEDFVE 240  
 Db 181 NALLRTPDLRWNSSNAQLLLDYCSSKGNYSWELGNPNPSFWKKAHILIDGLQGEDFVE 240  
 QY 241 LHKLLQSAFQNAKLYGPDIGPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300  
 Db 241 LHKLLQSAFQNAKLYGPDIGPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300  
 QY 301 DFLSSDALDTFILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDK 360  
 Db 301 DFLSSDALDTFILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDK 360  
 QY 361 LGLSAQMGIEVVMQVFFGAGNYHLVDENPEPLDYWLSSLFVKLVGPRVLLSRVKGPD 420  
 Db 361 LGLSAQMGIEVVMQVFFGAGNYHLVDENPEPLDYWLSSLFVKLVGPRVLLSRVKGPD 420

QY 421 SKLRVYLHCTNVYHPRYOEGDLTYVLNLHNVTYKHLKVPPLPRKPKVDYTLKPSGPDGL 480  
 Db |||||  
 QY 421 SKLRVYLHCTNVYHPRYOEGDLTYVLNLHNVTYKHLKVPPLPRKPKVDYTLKPSGPDGL 480  
 Db |||||  
 QY 481 LSKSVQLNGQILKXWVDEQTLPALTEKPLPAGSALSPLPAGSYGFFVIRNAKIAACI 535  
 Db |||||  
 QY 481 LSKSVQLNGQILKXWVDEQTLPALTEKPLPAGSALSPLPAGSYGFFVIRNAKIAACI 535  
 Db |||||

## RESULT 7

ADT78175  
 ID ADT78175 standard; protein; 535 AA.

AC ADT78175;  
 DT 13-JAN-2005 (first entry)  
 DE Mouse heparanase protein.

XX Antibody; epitope; heparanase; pathological condition; angiogenesis;  
 KW cell proliferation; cancerous condition; tumour cell invasion;  
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;  
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;  
 KW cytostatic; antiinflammatory; vulnery; antiarteriosclerotic;  
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; mouse.  
 XX

OS Mus musculus.

XX Key Location/Qualifiers

FT Binding-site 149..154

FT /note= "Putative heparin binding site"

FT Binding-site 263..269

FT /note= "Putative heparin binding site"

FT Binding-site 418..425

FT /note= "Putative heparin binding site"

XX US2004213789-A1.

XX 28-OCT-2004.

XX 22-AUG-2003; 2003US-00645659.

XX 02-SEP-1997; 97US-00922170.

XX 01-MAY-1998; 98US-00071739.

XX 04-NOV-1998; 98US-00186200.

XX 19-FEB-2003; 2003US-00368044.

XX (YACO/) YACOBY-ZEEVI O.

PA (PERE/) PERETZ T.

PA (MIRO/) MIRON D.

PA (SHLO/) SHLOMI Y.

PA (PECK/) PECKER I.

PA (AYAL/) AYAL-HERSHKOVITZ M.

PA (FEIN/) FEINSTEIN E.

PA (GELD/) GELDER J M V.

PA (VLOD/) VLODAVSKY I.

PA (FRIE/) FRIEDMANN Y.

XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;

PI Ayal-HersHKovitz M, Feinstein E, Gelder JMW, Vlodavsky I;

PI Friedmann Y;

XX WPI; 2004-774790/76.

XX New neutralizing monoclonal anti-heparanase antibodies, useful for

PT detecting, treating or preventing cancer, inflammatory or autoimmune

PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.

XX Claim 5; SEQ ID NO 2; 68pp; English.

XX The invention relates to an isolated antibody or antibody portion capable

XX of specifically binding to or elicited by at least one epitope of a

CC heparanase protein, where the heparanase protein is at least 60%  
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and  
 CC where at least one epitope comprises a sequence at least 70% homologous  
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)  
 CC a hybridoma cell line comprising a cell line for producing the monoclonal  
 CC antibody, (b) a method for detecting, treating or preventing a  
 CC pathological condition or a heparanase-related disorder or condition in a  
 CC subject, (c) a method for monitoring the state of a heparanase-related  
 CC disorder or condition in a subject, and (d) a pharmaceutical composition  
 CC comprising the isolated anti-heparanase antibody or antibody portion and  
 CC a pharmaceutical carrier. The antibody, methods, and composition are  
 CC useful for detecting, treating, preventing or monitoring a pathological  
 CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition  
 CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,  
 CC or prostate cancer), minor cell proliferation, invasion of circulating  
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder  
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy  
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or  
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,  
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell  
 CC carcinoma) in a mammal. This sequence represents mouse heparanase.  
 XX

XX Sequence 535 AA;

Query Match 100.0%; Score 2797; DB 8; Length 535;

Best Local Similarity 100.0%; Pred. No. 2.9e-262;

Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRLLLMLWGLGALAAQAGAPAGTAPDDVVDVDFYTKRPLRSVSPFLSTIDASLATD 60  
 Db |||||  
 QY 1 MLRLLLMLWGLGALAAQAGAPAGTAPDDVVDVDFYTKRPLRSVSPFLSTIDASLATD 60  
 Db |||||  
 QY 61 PRFLTFLGSPRLALARGLSPAYLRFQGTCTDFLIFDPDKPTSEERSYKWSQVNHICR 120  
 Db |||||  
 QY 61 PRFLTFLGSPRLALARGLSPAYLRFQGTCTDFLIFDPDKPTSEERSYKWSQVNHICR 120  
 Db |||||  
 QY 121 SEPVSAAVLRLQVWPFQELLLREQYQKEFKNSTYSRSDVMDLYSFAKCSGLDLIFGL 180  
 Db |||||  
 QY 121 SEPVSAAVLRLQVWPFQELLLREQYQKEFKNSTYSRSDVMDLYSFAKCSGLDLIFGL 180  
 Db |||||  
 QY 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDEVE 240  
 Db |||||  
 QY 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDEVE 240  
 Db |||||  
 QY 241 LHKLLQSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300  
 Db |||||  
 QY 241 LHKLLQSAFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300  
 Db |||||  
 QY 301 DFLSSDALDFTFILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDK 360  
 Db |||||  
 QY 301 DFLSSDALDFTFILSVQKILKVTKEITPGKKVWLGETSSAYGGGAPLLSNTFAAGFMWLDK 360  
 Db |||||  
 QY 361 LGLSAQMGIEVVMRQVFFGAGNYHLVDENFEPLPDYWLISLLFKLVGPRVLLSRVKGDR 420  
 Db |||||  
 QY 361 LGLSAQMGIEVVMRQVFFGAGNYHLVDENFEPLPDYWLISLLFKLVGPRVLLSRVKGDR 420  
 Db |||||  
 QY 421 SKLRVYLHCTNVYHPRYOEGDLTYVLNLHNVTYKHLKVPPLPRKPKVDYTLKPSGPDGL 480  
 Db |||||  
 QY 421 SKLRVYLHCTNVYHPRYOEGDLTYVLNLHNVTYKHLKVPPLPRKPKVDYTLKPSGPDGL 480  
 Db |||||  
 QY 481 LSKSVQLNGQILKXWVDEQTLPALTEKPLPAGSALSPLPAGSYGFFVIRNAKIAACI 535  
 Db |||||  
 QY 481 LSKSVQLNGQILKXWVDEQTLPALTEKPLPAGSALSPLPAGSYGFFVIRNAKIAACI 535  
 Db |||||

## RESULT 8

AEA42424

ID AEA42424 standard; protein; 535 AA.

XX AEA42424;

XX AEA42424;

XX 28-JUL-2005 (first entry)

XX

Mouse heparanase epitope peptide SEQ ID NO:2.

antibody; heparanase; antiinflammatory; vulnary; immunosuppressive;  
 antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;  
 inflammation; wound healing; scarring; vasculopathy; autoimmune disease;  
 angiogenesis disorder; cancer; tumor; metastasis.

OS Mus musculus.

FN AU2004201462-A1.

XX PD 06-MAY-2004.

XX PF 08-APR-2004; 2004AU-00201462.

XX PR 08-APR-2004; 2004AU-00201462.

XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.

XX Vlodavsky I, Pecker I, Miron M, Gilboa A, Miron D, Moskowitz H;  
 PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-HersHKovitz M, Ben-Artzi H;  
 PI Feinstein E;

XX WPI; 2005-173343/19.

XX Novel isolated antibody capable of specifically binding to epitope of  
 PT heparanase protein, useful for preventing and treating heparanase-related  
 PT disorder such as inflammatory disorder, scars, autoimmune conditions or  
 XX angiogenesis.

PS Claim 2; SEQ ID NO 2; 260pp; English.

XX The invention relates to an isolated antibody or its portion (I) capable  
 CC of specifically binding to an epitope of a heparanase protein. Also  
 CC described: (1) a cell line (II) for producing a monoclonal antibody or  
 CC its portion, comprising a cell line for producing (I); (2) a  
 CC pharmaceutical composition comprising (I) and a carrier; and (3) an  
 CC affinity medium (III) for binding human heparanase polypeptides,  
 CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)  
 CC useful for treating a subject suffering from a pathological condition,  
 CC which involves administering (I) to the subject. (I) is useful for  
 CC preventing and treating heparanase-related disorder or condition chosen  
 CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune  
 CC condition, angiogenesis, cell proliferation, cancerous condition, tumor  
 CC cell proliferation, invasion of circulating tumor cells and metastatic  
 CC disease. (I) is useful for detecting the presence of heparanase  
 CC polypeptide in a sample. (I) is useful for detecting heparanase-related  
 CC disease or condition in a subject such as vertebrate, preferably mammal  
 CC e.g., human. The heparanase-related disorder or condition further  
 CC includes renal disease or disorder chosen from diabetic nephropathy,  
 CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome  
 CC and renal cell carcinoma. The present sequence represents mouse  
 CC heparanase, which is used in the exemplification of the present  
 CC invention.

XX Sequence 535 AA;

Query Match 100.0%; Score 2797; DB 9; Length 535;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-262;  
 Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLRLLLLLWGPGLGALQAQAGTAPTDVDDVDFEYTKRLRSVSPSFLSITTDASLATD 60  
 DB 1 MLRLLLLLWGPGLGALQAQAGTAPTDVDDVDFEYTKRLRSVSPSFLSITTDASLATD 60

QY 61 PRELTLGSPRLRALARGLSPAYLRFGGTKTDFLIFDPDKPTSEERSYKWSQVNHDIICR 120  
 DB 61 PRELTLGSPRLRALARGLSPAYLRFGGTKTDFLIFDPDKPTSEERSYKWSQVNHDIICR 120

QY 121 SEPVSAAVLKQLQVWPFQBELLLREYQKEFKNSTYSRSSVDMLYSFACSGLDLIFGL 180  
 DB 121 SEPVSAAVLKQLQVWPFQBELLLREYQKEFKNSTYSRSSVDMLYSFACSGLDLIFGL 180

QY 181 NALLRTPDLRNWNSSNAQLLLDYCSSKGYNIWELGNEPNSFWKKAHILIDGLQGEDFVE 240  
 DB 181 NALLRTPDLRNWNSSNAQLLLDYCSSKGYNIWELGNEPNSFWKKAHILIDGLQGEDFVE 240

QY 241 LHKLQSAFQNAKLYGPDICQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300  
 DB 241 LHKLQSAFQNAKLYGPDICQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300

QY 301 DFLSSDALDTPILSVQKILKVTKEITPOKKVWLGTSSTAYGGGAPLLSNTFAAGFWLDDK 360  
 DB 301 DFLSSDALDTPILSVQKILKVTKEITPOKKVWLGTSSTAYGGGAPLLSNTFAAGFWLDDK 360

QY 361 LGLSAQMGIEVVMROVFFGAGNYHLVDENFPLPDYWLSSLFVKLVGRVLLSRVKGPDOR 420  
 DB 361 LGLSAQMGIEVVMROVFFGAGNYHLVDENFPLPDYWLSSLFVKLVGRVLLSRVKGPDOR 420

QY 421 SKLRVYLHCTNVYHPRYQEGDLTLVNLHNVTKLKVPPPLFRKPVDVDTYLLKPSGPDGL 480  
 DB 421 SKLRVYLHCTNVYHPRYQEGDLTLVNLHNVTKLKVPPPLFRKPVDVDTYLLKPSGPDGL 480

QY 481 LSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSPLAFSGFFVIRNAKIAACI 535  
 DB 481 LSKSVQLNGQILKMWDEQTLPALTEKPLPAGSALSPLAFSGFFVIRNAKIAACI 535

RESULT 9  
 ADY27033  
 ID ADY27033 standard; protein; 535 AA.

XX AC ADY27033;

XX DT 05-MAY-2005 (first entry)

XX DE Murine heparanase protein.

XX KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
 KW neurological disease; viral infection; infection; cytostatic;  
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
 KW protease; enzyme; enzyme purification.

XX OS Mus musculus.

XX FN WO2005016227-A2.

XX PD 24-FEB-2005.

XX PF 12-AUG-2004; 2004WO-IL000744.

XX PR 14-AUG-2003; 2003US-0494800P.

XX PR 12-JAN-2004; 2004US-0535492P.

XX PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX PI Van-Gelder JM, Miron D;

XX DR WPI; 2005-182203/19.

XX PT Regulating heparanase activity, useful for treating heparanase-associated  
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases, heparanase  
 PT neurological diseases or viral diseases) comprises modulating heparanase  
 PT activation.

XX PS Disclosure; SEQ ID NO 5; 211pp; English.

XX CC The invention relates to a method of regulating heparanase activity in a  
 CC tissue or regulating a biological process depending at least in part on  
 CC heparanase activity comprising modulating heparanase activation. The  
 CC invention also relates to methods of treating a heparanase- or heparin  
 CC binding protein-associated disease or disorder in a subject, a  
 CC pharmaceutical composition for use in the treatment of a heparanase-  
 CC associated disease or disorder comprising a therapeutic amount of an  
 CC agent capable of modulating heparanase activation and a pharmaceutical

CC carrier or diluent, a method of identifying a protease activator of  
CC heparanase, a protease substrate mimetic comprising a peptide  
CC representing a subset or all substrate residues or cleavage sites of  
CC human heparanase or an equivalent non-human heparanase, a method of  
CC producing active heparanase and a method of modulating an adhesion  
CC activity of heparanase. The composition and methods are useful for  
CC modulating heparanase activation and for treating heparanase-associated  
CC diseases or disorders such as cancer, inflammation, cardiovascular  
CC diseases, neurological diseases or viral infections. This sequence  
CC represents a murine heparanase protein used in the scope of the  
CC invention.  
XX  
XX Sequence 535 AA;  
SQ  
  
Query Match 99.9%; Score 2793; DB 9; Length 535;  
Best Local Similarity 99.8%; Pred. No. 7.2e-262;  
Matches 534; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 MLRLLLWLPGLAQAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPFLSITIDASLATD 60  
DB 1 MLRLLLWLPGLAQAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPFLSITIDASLATD 60  
  
QY 61 PRFLTFLGSPRLRALARGLSPAYLRFGGTTDFLI FDPDKPTSEERSYKWSQVNHDI 120  
DB 61 PRFLTFLGSPRLRALARGLSPAYLRFGGTTDFLI FDPDKPTSEERSYKWSQVNHDI 120  
  
QY 121 SEPVSAAVLRLKQVWPFQELLRLRQYQKEFNSTYSRSSVDMLYSFAKCSGLDLIFGL 180  
DB 121 SEPVSAAVLRLKQVWPFQELLRLRQYQKEFNSTYSRSSVDMLYSFAKCSGLDLIFGL 180  
  
QY 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDFVE 240  
DB 181 NALLRTPDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDFVE 240  
  
QY 241 LHKLLQRSFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300  
DB 241 LHKLLQRSFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATKE 300  
  
QY 301 DFLSSDALDTFILSVOKILKVTKEITPGKKVWLGESSTAYGGGAPLLSNTFAAGFMWLDK 360  
DB 301 DFLSSDVLTDFILSVOKILKVTKEITPGKKVWLGESSTAYGGGAPLLSNTFAAGFMWLDK 360  
  
QY 361 LGLSAQMIEVNRQVFFGAGNHYLVNDFENPEPLDYWLSLLFKLGVPRVLLSRVKGPD 420  
DB 361 LGLSAQMIEVNRQVFFGAGNHYLVNDFENPEPLDYWLSLLFKLGVPRVLLSRVKGPD 420  
  
QY 421 SKLRVYLHCTNVYHPRYQEGDLTYLVNLHNVTKHLKVPDPPLFRKPVDTYLLKPSGPDGL 480  
DB 421 SKLRVYLHCTNVYHPRYQEGDLTYLVNLHNVTKHLKVPDPPLFRKPVDTYLLKPSGPDGL 480  
  
QY 481 LSKSVOLNGOILKMWDEQTLPALTEKPLPAGSALSIPAFSGFFVIRNAKIAACI 535  
DB 481 LSKSVOLNGOILKMWDEQTLPALTEKPLPAGSALSIPAFSGFFVIRNAKIAACI 535

RESULT 10  
ABB07812  
ID ABB07812 standard; protein; 536 AA.  
XX  
AC ABB07812;  
XX  
DT 03-JUL-2002 (first entry)  
XX  
DE Rat heparanase sequence.  
XX  
KW Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;  
KW anti-protozoan; neuroprotective; heparin; rat.  
XX  
OS Rattus rattus.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..16  
FT /note= "putative signal peptide"  
FT  
FT

FT Protein 17..536  
FT /note= "mature protein"  
XX  
PN US2002034810-A1.  
XX  
PD 21-MAR-2002.  
XX  
PF 16-AUG-2001; 2001US-00930218.  
XX  
PR 20-SEP-2000; 2000US-00666390.  
XX  
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
XX  
XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia B;  
XX WPI; 2002-338926/37.  
XX  
XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful  
PT to treat various heparin-related disorders and the signal peptide is  
PT useful in production of membrane-targeted or secreted recombinant  
PT proteins.  
XX  
PS Disclosure; Fig 1a; 39pp; English.  
XX  
XX The invention relates to an isolated avian and reptile nucleic acid,  
CC encoding a polypeptide with heparanase catalytic activity. The signal  
CC peptide of the nucleic acid can be used to express membrane-associated or  
CC secreted proteins in heterologous expression systems. The encoded  
CC polypeptides can be used to prevent tumour angiogenesis, metastasis and  
CC invasion, and to intervene with pathologies associated with impaired  
CC heparin-binding growth factors, cellular responses to heparin-binding  
CC growth factors and cytokines, cell interaction with plasma lipoproteins,  
CC cellular susceptibility to viral, protozoa and bacterial infections or  
CC disintegration of neurodegenerative plaques. The present sequence  
CC represents a rat heparanase protein sequence used in similarity studies  
XX  
XX Sequence 536 AA;  
SQ  
  
Query Match 92.6%; Score 2590.5; DB 5; Length 536;  
Best Local Similarity 92.5%; Pred. No. 3.5e-242;  
Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;  
  
QY 1 MLR-LLLLLWLPGLAQAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPFLSITIDASLAT 59  
DB 1 MLRPLLLLWLPGLAQAQAGAPAGTAPDDVVDLEFYTKRPLRSVSPFLSITIDASLAT 60  
  
QY 60 DPRFLTFLGSPRLRALARGLSPAYLRFGGTTDFLI FDPDKPTSEERSYKWSQVNHDI 119  
DB 61 DPRFLTFLGSPRLRALARGLSPAYLRFGGTTDFLI FDPDKPTSEERSYKWSQVNHDI 120  
  
QY 120 RSEPVSAAVLRKQVWPFQELLRLRQYQKEFNSTYSRSSVDMLYSFAKCSGLDLIFG 179  
DB 121 GSERVSADVLRKQVWPFQELLRLRQYQKEFNSTYSRSSVDMLYSFAKCSGLDLIFG 180  
  
QY 180 LNALRTPDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHILIDGLQGEDFV 239  
DB 181 LNALRTPDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAQISIDGLQGEDFV 240  
  
QY 240 ELHKLLQRSFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRIATK 299  
DB 241 ELHKLLQRSFQNAKLYGPDIGQPRGKTVKLLRSFLKAGGEVIDSLTWHYYLNGRVATK 300  
  
QY 300 EDFLSSDALDTFILSVOKILKVTKEITPGKKVWLGESSTAYGGGAPLLSNTFAAGFMWLD 359  
DB 301 EDFLSSDVLTDFILSVOKILKVTKEITPGKKVWLGESSTAYGGGAPLLSNTFAAGFMWLD 360  
  
QY 360 KLGLSAQMIEVNRQVFFGAGNHYLVNDFENPEPLDYWLSLLFKLGVPRVLLSRVKGPD 419  
DB 361 KLGLSAQMIEVNRQVFFGAGNHYLVNDFENPEPLDYWLSLLFKLGVPRVLLSRVKGPD 420  
  
QY 420 RSKLRVYLHCTNVYHPRYQEGDLTYLVNLHNVTKHLKVPDPPLFRKPVDTYLLKPSGPDG 479  
DB 421 RSKLRVYLHCTNVYHPRYQEGDLTYLVNLHNVTKHLKVPDPPLFRKPVDTYLLKPSGPDG 480







XX ADT78176;  
 AC 13-JAN-2005 (first entry)  
 DT Rat heparanase protein.  
 DE  
 XX Antibody; epitope; heparanase; pathological condition; angiogenesis;  
 KW cell proliferation; cancerous condition; tumour cell invasion;  
 KW metastatic disease; heparanase-related disorder; inflammatory disorder;  
 KW wound; scar; vasculopathy; autoimmune condition; renal disease;  
 KW cytostatic; antiinflammatory; vulnerable; antiarteriosclerotic;  
 KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; rat.  
 XX Rattus norvegicus.  
 OS  
 XX Key Location/Qualifiers  
 FH Binding-site 150..155  
 FT /note= "Putative heparin binding site"  
 FT Binding-site 264..270  
 FT /note= "Putative heparin binding site"  
 FT Binding-site 419..426  
 FT /note= "Putative heparin binding site"  
 XX US2004213789-A1.  
 PN 28-OCT-2004.  
 PD  
 XX 22-AUG-2003; 2003US-00645659.  
 XX  
 XX 02-SEP-1997; 97US-00922170.  
 PR 01-MAY-1998; 98US-00071739.  
 PR 04-NOV-1998; 98US-00186200.  
 PR 19-FEB-2003; 2003US-00368044.  
 XX  
 XX (YACO/) YACOBY-ZEEVI O.  
 PA (PERE/) PERETZ T.  
 PA (MIRO/) MIRON D.  
 PA (SHLO/) SHLOMI Y.  
 PA (PECK/) PECKER I.  
 PA (AYAL/) AYAL-HERSHKOVITZ M.  
 PA (FEIN/) FEINSTEIN E.  
 PA (GELD/) GELDER J M V.  
 PA (VLOD/) VLODAVSKY I.  
 PA (FRIE/) FRIEDMANN Y.  
 XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
 PI Ayal-Hershkovitz M, Feinstein E, Gelder JMV, Vlodavsky I;  
 PI Friedmann Y;  
 XX WPI; 2004-774790/76.  
 XX  
 XX New neutralizing monoclonal anti-heparanase antibodies, useful for  
 PT detecting, treating or preventing cancer, inflammation or autoimmune  
 PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.  
 PS  
 XX Claim 5; SEQ ID NO 3; 68pp; English.  
 XX  
 XX The invention relates to an isolated antibody or antibody portion capable  
 CC of specifically binding to or elicited by at least one epitope of a  
 CC heparanase protein, where the heparanase protein is at least 60%  
 CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and  
 CC where at least one epitope comprises a sequence of at least 70% homologous  
 CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)  
 CC a hybridoma cell line comprising a cell line for producing the monoclonal  
 CC antibody, (b) a method for detecting, treating or preventing a  
 CC pathological condition or a heparanase-related disorder or condition in a  
 CC subject, (c) a method for monitoring the state of a heparanase-related  
 CC disorder or condition in a subject, and (d) a pharmaceutical composition  
 CC comprising the isolated anti-heparanase antibody or antibody portion and  
 CC a pharmaceutical carrier. The antibody, methods, and composition are  
 CC useful for detecting, treating, preventing or monitoring a pathological  
 CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition

CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,  
 CC or prostate cancer), minor cell proliferation, invasion of circulating  
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder  
 CC on condition, e.g. an inflammatory disorder, wound, scar, vasculopathy  
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or  
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,  
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell  
 CC carcinoma) in a mammal. This sequence represents rat heparanase.  
 XX  
 SQ Sequence 536 AA;  
 Query Match 92.6%; Score 2590.5; DB 8; Length 536;  
 Best Local Similarity 92.5%; Pred. No. 3.5e-242;  
 Matches 496; Conservative 19; Mismatches 20; Indels 1; Gaps 1;  
 QY 1 MLR-LLLMLWGPGLGALAQGAPAGTAPTDVVDLEFYTTPRPLRSVSPFLSITIDASLAT 59  
 DB 1 MLRPLLLLMLWGRRLALTQGTAGTAPTKDQVDLEFYTTPRPLRSVSPFLSITIDASLAT 60  
 QY 60 DPRFLTFLGSPRLRALARGLSPAYLRFEGGTTDFLIFDPDKPTSEERSYWKSVQNHQD 119  
 DB 61 DPRFLTFLGSPRLRALARGLSPAYLRFEGGTTDFLIFDPDKPTSEERSYWKSVQNHQD 120  
 QY 120 RSEPVSAVLRLQVWPPFQELLRLRLRLRLRLRLRLRLRLRLRLRLRLRLRLRLRL 179  
 DB 121 GSERVSADVLRLQVWPPFQELLRLRLRLRLRLRLRLRLRLRLRLRLRLRLRLRL 180  
 QY 180 LNALRTPLDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHTLIDGLQGEDFV 239  
 DB 181 LNALRTPLDLRWNSNAQLLDYCSSKGYNISWELGNEPNSFWKKAHTLIDGLQGEDFV 240  
 QY 240 ELHKLQSAFONAKLYGPDIGQPRGKTIVKLLRSFLKAGGEVIDSLTWHHYLLNGRIATK 299  
 DB 241 ELHKLQSAFONAKLYGPDIGQPRGKTIVKLLRSFLKAGGEVIDSLTWHHYLLNGRVATK 300  
 QY 300 EDFLSSDALDTFILSVQIKLVTKETTPGKKVWLGGETSSAYGGGAPLLSNTFAAGFMWLD 359  
 DB 301 EDFLSSDALDTFILSVQIKLVTKETTPGKKVWLGGETSSAYGGGAPLLSNTFAAGFMWLD 360  
 QY 360 KLGLSAQMGLEVVMQVFFCGAGNVLVDENFELPDYWLSLFLKLVGPRVLLSRVKGPD 419  
 DB 361 KLGLSAQMGLEVVMQVFFCGAGNVLVDENFELPDYWLSLFLKLVGPRVLLSRVKGPD 420  
 QY 420 RSKLRVYLHCTNVYHPRYOEGDLTYVLNLHNVTKHLKVPPLPKPVPDVTLLKPSGPDG 479  
 DB 421 RSKLRVYLHCTNVYHPRYOEGDLTYVLNLHNVTKHLKVPPLPKPVPDVTLLKPSGPDG 480  
 QY 480 LLSKSVQLNGQILKMVDEQTLPALETEKPLPAGSALSIPAFSYGFFVIRNAKIAACI 535  
 DB 481 LLSKSVQLNGQILKMVDEQTLPALETEKPLPAGSALSIPAFSYGFFVIRNAKIAACI 536  
 RESULT 13  
 ADY27035  
 ID ADY27035 standard; protein; 536 AA.  
 XX  
 AC ADY27035;  
 XX  
 DT 05-MAY-2005 (first entry)  
 DE Rat heparanase protein.  
 XX  
 KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
 KW neurological disease; viral infection; infection; cytostatic;  
 KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
 XX protease; enzyme; endyme purification.  
 OS Rattus norvegicus.  
 XX  
 PN W02005016227-A2.  
 XX  
 PD 24-FEB-2005.  
 XX

Pf	12-AUG-2004; 2004WO-IL000744.
Xx	14-AUG-2003; 2003US-0494800P.
Pf	12-JAN-2004; 2004US-0535492P.
Xx	(INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
Xx	Van-Gelder JM, Miron D;
Pt	WPI; 2005-182203/19.
Xx	
Xx	Regulating heparanase activity, useful for treating heparanase-associated
Pt	diseases (e.g. cancer, inflammation, cardiovascular diseases,
Pt	neurological diseases or viral diseases) comprises modulating heparanase
Pt	activation.
Xx	
Pt	Disclosure; SEQ ID NO 7; 211pp; English.
Pt	
Xx	The invention relates to a method of regulating heparanase activity in a
Cc	tissue or regulating a biological process depending at least in part on
Cc	heparanase activity comprising modulating heparanase activation. The
Cc	invention also relates to methods of treating a heparanase- or heparin
Cc	binding protein-associated disease or disorder in a subject, a
Cc	pharmaceutical composition for use in the treatment of a heparanase-
Cc	associated disease or disorder comprising a therapeutic amount of an
Cc	agent capable of modulating heparanase activation and a pharmaceutical
Cc	carrier or diluent, a method of identifying a protease activator of
Cc	heparanase, a protease substrate mimetic comprising a peptide
Cc	representing a subset or all substrate residues or cleavage sites of
Cc	human heparanase or an equivalent non-human heparanase, a method of
Cc	producing active heparanase and a method of modulating an adhesion
Cc	activity of heparanase. The composition and methods are useful for
Cc	modulating heparanase activation and for treating heparanase-associated
Cc	diseases or disorders such as cancer, inflammation, cardiovascular
Cc	diseases, neurological diseases or viral infections. This sequence
Cc	represents a rat heparanase protein used in the scope of the invention.
Xx	
Sq	Sequence 536 AA;
	Query Match            92.6%; Score 2590.5; DB 9; Length 536;
	Best Local Similarity   92.5%; Pred. No. 3.5e-242; Indels   1; Gaps   1;
	Matches 496; Conservative   19; Mismatches   20;
Qy	1 MLR-LLLLWLWGPLGALAAQAAGACTAPTDDVVLDLEFYTKRLPSVSPSLITIDASLAT 59
Dd	1 MLRPLLLLLWGLRALTLTGTFPACTPTKDVVDLEFYTKRLFQSVSPSLITIDASLAT 60
Qy	60 DRPFLTPLGSPRLARLARGLSPAYLRFGTGTDFLIIFDPDKEPTSEERSYWKSVQHDIIC 119
Dd	61 DRPFLTPLGSPRLARLARGLSPAYLRFGTGTDFLIIFDPNKEPTSEERSYMQSDNDNIC 120
Qy	120 RSRPVSAVLRLQLVEMPFQELLLLRQYOKEFKNSTYSRSSVDMLVSYFACSGLDLIFG 179
Dd	121 GSERVSADVLRKLQMEFPFQELLLLRQYQREFKNSTYSRSSVDMLVSYFACSRDLIFG 180
Qy	180 LNALLRTPDLRWNSNAQLLLDYCSKGYNIISWELGNPNFSFWKAHLIIDGLQGEDVF 239
Dd	181 LNALLRTPDLRWNSNAQLLLNYCCKGYNISWELGNPNFSFWKAKQISIDLQLGEDVF 240
Qy	240 ELHKLLORSAFONAKLYGPDIGQRGKTVKLLRSFLKAGGBVI DSLTWHHYYLNGRIATK 299
Dd	241 ELHKLLQSFAONAKLYGPDIGQRGKTVKLLRSFLKAGGBVI DSLTWHHYYLNGRVATK 300
Qy	300 EDFLSSDALDTFILSVOKILKVTEITPGKKVMJGETSSAYGGGAPLLSNTPAAGFMWLD 359
Dd	301 EDFLSSDVLDTFILSVOKILKVTEITPGKKVMJGETSSAYGGGAPLLSNTPAAGFMWLD 360
Qy	360 KLGLSQMGI EVVMRQVFFGAGNYHLVDENFEPLPDYWLSSLFKLVGPRLVRKVGPD 419
Dd	361 KLGLSQLGLIEVVMRQVFFGAGNYHLVDENFEPLPDYWLSSLFKLVGPRLVRKVGPD 420
Qy	420 RSKLRVYLCTNVYHPYQSGDLFTLYVLNLHNVTGLKVPPLPRKPVDVTVLLKPSGDG 479

421	RSKURVYLHCTNVVPHRYREGDFTLVNLNHNVTXKLLPDPMPFRSPVDVXKLLKPFSGDG	480
480	LLSKSVQLNGQILKQWVDEQTLPALTEKPLPAGSSALSLVPAFSYGFFVIRNAKIAACI	535
481	LLSKSVQLNGQTLKQWVDEQTLPALTEKPLPAGSSLSVPAFSYGFFVIRNAKIAACI	536
Db		
Qy		
Db		
RESULT 14		
AEA42425		
ID	AEA42425 standard; protein; 536 AA.	
XX		
AC	AEA42425;	
XX		
XX	28-JUL-2005 (first entry)	
XX		
DE	Rat heparanase epitope peptide SEQ ID NO:3.	
XX		
KW	antibody; heparanase; antiinflammatory; vulnerary; immunosuppressive;	
KW	antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;	
KW	inflammation; wound healing; scarring; vasculopathy; autoimmune disease;	
KW	angiogenesis disorder; cancer; tumor; metastasis.	
XX		
OS	Rattus norvegicus.	
XX		
PN	AU2004201462-A1.	
XX		
PD	06-MAY-2004.	
XX		
PF	08-APR-2004; 2004AU-00201462.	
XX		
PR	08-APR-2004; 2004AU-00201462.	
XX		
PA	(INST-) INSIGHT BIOPHARMACEUTICALS LTD.	
PA	(HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.	
XX		
XX	Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;	
PI	Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-HersHKovitz M, Ben-Artzi H;	
PI	Feinstein E;	
XX		
DR	WPI; 2005-173343/19.	
XX		
PT	Novel isolated antibody capable of specifically binding to epitope of	
PT	heparanase protein, useful for preventing and treating heparanase-related	
PT	disorder such as inflammatory disorder, scars, autoimmune conditions or	
PT	angiogenesis.	
XX		
PS	Claim 2; SEQ ID NO 3; 260pp; English.	
XX		
CC	The invention relates to an isolated antibody or its portion (I) capable	
CC	of specifically binding to an epitope of a heparanase protein. Also	
CC	described: (1) a cell line (II) for producing a monoclonal antibody or	
CC	its portion, comprising a cell line for producing (I); (2) a	
CC	pharmaceutical composition comprising (I) and a carrier; and (3) an	
CC	affinity medium (III) for binding human heparanase polypeptides,	
CC	comprising (I) immobilized to a chemically inert, insoluble carrier. (I)	
CC	useful for treating a subject suffering from a pathological condition,	
CC	which involves administering (I) to the subject. (I) is useful for	
CC	preventing and treating heparanase-related disorder or condition chosen	
CC	from inflammatory disorder, wound, scar, vasculopathy, autoimmune	
CC	condition, angiogenesis, cell proliferation, cancerous condition, tumor	
CC	cell proliferation, invasion of circulating tumor cells and metastatic	
CC	disease. (I) is useful for detecting the presence of heparanase	
CC	polypeptide in a sample. (I) is useful for detecting heparanase-related	
CC	disease or condition in a subject such as vertebrate, preferably mammal	
CC	e.g., human. The heparanase-related disorder or condition further	
CC	includes renal disease or disorder chosen from diabetic nephropathy,	
CC	glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome,	
CC	and renal cell carcinoma. The present sequence represents rat heparanase,	
CC	which is used in the exemplification of the present invention.	
XX		
SQ	Sequence 536 AA;	

Query Match 92.6%; Score 2590.5; DB 9; Length 536;

Best Local Similarity	92.5%;	Pred. No. 3.5e-242;			
Matches	496;	Conservative	19;	Mismatches	20;
				Indels	1;
				Gaps	1;
QY	1	MLR-LLLLMWGLGALAAQAGAPAGTAPD	VDVLEFYTKRPLRSVSPFLSITIDASLAT	59	
DB	1	MLRPLLLLWGLRLALTCQTPAGTAPKDV	DVLEFYTKRPLRSVSPFLSITIDASLAT	60	
QY	60	DRPFLFLGSPRLALARGLSPAYLRFGG	TKTDFLIDPDKEPTSEERSYKQVNHDC	119	
DB	61	DRPFLFLGSPRLALARGLSPAYLRFGG	TKTDFLIDPDKEPTSEERSYKQVNHDC	120	
QY	120	RSPVSAALVRKLQVWPFQELLRLREOY	KFNSTYSRSDVMLYSFAKCSGLDLIFG	179	
DB	121	GSRVSADVLRLQVWPFQELLRLREOY	KFNSTYSRSDVMLYSFAKCSGLDLIFG	180	
QY	180	LNALLTPDLRWNSNAQLLLDYCSSKGY	NIWELGNEPNSFWKKAHILIDGLQGD	239	
DB	181	LNALLTPDLRWNSNAQLLLDYCSSKGY	NIWELGNEPNSFWKKAHILIDGLQGD	240	
QY	240	ELHKLQSAFQNAKLYGPDIGQPRGKT	VKLLRSFLKAGGEVIDSLTWHHYLNG	299	
DB	241	ELHKLQSAFQNAKLYGPDIGQPRGKT	VKLLRSFLKAGGEVIDSLTWHHYLNG	300	
QY	300	EDPLSSDALDTFLSVQKILKVTKEIT	PPGKKVWLGETSSAYGGAPLLSNTFA	359	
DB	301	EDPLSSDALDTFLSVQKILKVTKEIT	PPGKKVWLGETSSAYGGAPLLSNTFA	360	
QY	360	KLGLSAQMGIEVVMRQVFFGAGNYHL	VNDENFEPLPDYWLSSLFKLVGPRV	419	
DB	361	KLGLSAQMGIEVVMRQVFFGAGNYHL	VNDENFEPLPDYWLSSLFKLVGPRV	420	
QY	420	RSKRLVYLHCTNVYHPRYQEGDITL	YVLNLHNTYKHLKVPPLFRKPDV	479	
DB	421	RSKRLVYLHCTNVYHPRYQEGDITL	YVLNLHNTYKHLKVPPLFRKPDV	480	
QY	480	LLSKSVQNLQGLKQVDEQTLPALTE	KPLPAGSALSIPAFSYGFFVIRNAK	535	
DB	481	LLSKSVQNLQGLKQVDEQTLPALTE	KPLPAGSALSIPAFSYGFFVIRNAK	536	
RESULT 15					
ADY27034	standard; protein; 545 AA.				
ID	ADY27034				
AC	ADY27034;				
XX	ADY27034;				
DT	05-MAY-2005 (first entry)				
XX	Bovine heparanase protein.				
XX	Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;				
KW	neurological disease; viral infection; infection; cytostatic;				
KW	antiflammatory; cardiovascular-gen.; neuroprotective; virucide;				
XX	protease; enzyme; enzyme purification.				
OS	Bos taurus.				
XX	WO2005016227-A2.				
PN	24-FEB-2005.				
XX	12-AUG-2004; 2004WO-IL000744.				
XX	14-AUG-2003; 2003US-0494800P.				
PR	12-JAN-2004; 2004US-0535492P.				
XX	(INSI-) INSIGHT BIOPHARMACEUTICALS LTD.				
PA	Van-Gelder JM, Miron D;				
PI	WPI; 2005-182203/19.				
XX	Regulating heparanase activity, useful for treating heparanase-associated				

PT	diseases (e.g. cancer, inflammation, cardiovascular diseases,				
PT	neurological diseases or viral diseases) comprises modulating heparanase				
XX	activation.				
PS	Disclosure; SEQ ID NO 6; 211pp; English.				
XX	The invention relates to a method of regulating heparanase activity in a				
CC	tissue or regulating a biological process depending at least in part on				
CC	heparanase activity comprising modulating heparanase activation. The				
CC	invention also relates to methods of treating a heparanase- or heparin				
CC	binding protein-associated disease or disorder in a subject, a				
CC	pharmaceutical composition for use in the treatment of a heparanase-				
CC	associated disease or disorder comprising a therapeutic amount of an				
CC	agent capable of modulating heparanase activation and a pharmaceutical				
CC	carrier or diluent, a method of identifying a protease activator of				
CC	heparanase, a protease substrate mimetic comprising a peptide				
CC	representing a subset or all substrate residues or cleavage sites of				
CC	human heparanase or an equivalent non-human heparanase, a method of				
CC	producing active heparanase and a method of modulating an adhesion				
CC	activity of heparanase. The composition and methods are useful for				
CC	modulating heparanase activation and for treating heparanase-associated				
CC	diseases or disorders such as cancer, inflammation, cardiovascular				
CC	diseases, neurological diseases or viral infections. This sequence				
CC	represents a bovine heparanase protein used in the scope of the				
XX	invention.				
XX	Sequence 545 AA;				
QY	Query Match	78.1%;	Score 2184;	DB 9;	Length 545;
DB	Best Local Similarity	77.8%;	Pred. No. 1.2e-202;		
DB	Matches	414;	Conservative	46;	Mismatches 72; Indels 0; Gaps 0;
QY	4	LLLLLWGLGALAAQAGAPAGTAPD	VDVLEFYTKRPLRSVSPFLSITIDASLAT	DRF 63	
DB	14	LLLLLWGLGALAAQAGAPAGTAPD	VDVLEFYTKRPLRSVSPFLSITIDASLAT	DRF 73	
QY	64	LTFGLSPRLALARGLSPAYLRFGG	TKTDFLIDPDKEPTSEERSYKQVNHDC	RSPE 123	
DB	74	FTFLGSKLTARGLAPAYLRFGG	TKTDFLIDPDKEPTSEERSYKQVNHDC	RSPE 133	
QY	124	VSAVLRKLQVWPFQELLRLREOY	KFNSTYSRSDVMLYSFAKCSGLDLIFG	NAL 183	
DB	134	IPSDVEEKLQVWPFQELLRLREOY	KFNSTYSRSDVMLYSFAKCSGLDLIFG	NAL 193	
QY	184	LRTPLDRWNSNAQLLLDYCSSKGY	NIWELGNEPNSFWKKAHILIDGLQGD	FVELHK 243	
DB	194	LRTPLDRWNSNAQLLLDYCSSKGY	NIWELGNEPNSFWKKAHILIDGLQGD	FVELHK 253	
QY	244	LLQSAFQNAKLYGPDIGQPRGKT	VKLLRSFLKAGGEVIDSLTWHHYLNG	RIATKEDFL 303	
DB	254	LLQSAFQNAKLYGPDIGQPRGKT	VKLLRSFLKAGGEVIDSLTWHHYLNG	RIATKEDFL 313	
QY	304	SSDALDTFLSVQKILKVTKEIT	PPGKKVWLGETSSAYGGAPLLSNTFA	AGFMWLDKGL 363	
DB	314	NPDILDTFLSVQKILKVTKEIT	PPGKKVWLGETSSAYGGAPLLSNTFA	AGFMWLDKGL 373	
QY	364	SAQMGIEVVMRQVFFGAGNYHL	VNDENFEPLPDYWLSSLFKLVGPRV	LLSVKVPDRSKL 423	
DB	374	SARMGIEVVMRQVFFGAGNYHL	VNDENFEPLPDYWLSSLFKLVGPRV	LLSVKVPDRSKL 433	
QY	424	RVYLHCTNVYHPRYQEGDITL	YVLNLHNTYKHLKVPPLFRKPDV	TYLLKPSGDL 483	
DB	434	RVYLHCTNVYHPRYQEGDITL	YVLNLHNTYKHLKVPPLFRKPDV	TYLLKPSGDL 493	
QY	484	SVQNLQGLKQVDEQTLPALTE	KPLPAGSALSIPAFSYGFFVIRNAK	IAACI 535	
DB	494	SVQNLQGLKQVDEQTLPALTE	KPLPAGSALSIPAFSYGFFVIRNAK	IAACI 545	

Search completed: June 5, 2006, 12:09:44  
Job time : 107.661 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:21 ; Search time 95.7733 Seconds  
(without alignments)  
3728.138 Million cell updates/sec

Title: US-10-645-659A-1  
Perfect score: 2020  
Sequence: 1 KFFKSTYSRSSVDVLYTFA.....LPAFSYSFVIRNAKVAACI 386

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 7.2.1\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2020	100.0	543	1	HPSE HUMAN
2	1724	85.3	574	2	Q333x9 _PRODE
3	1723	85.3	574	2	Q333x8 _PRODE
4	1711	84.7	574	2	Q333x7 _PRODE
5	1710	84.7	574	2	Q333x6 _SPAJD
6	1681	83.2	545	1	HPSE BOVIN
7	1661	82.2	535	1	HPSE MOUSE
8	1646	79.3	536	1	HPSE RAT
9	1602	79.3	558	2	Q333x5 _SPAJD
10	1345.5	66.6	523	1	HPSE CHICK
11	1059.5	52.5	533	2	Q4SYF6 _TETNG
12	978	48.4	592	1	HPSE2 HUMAN
13	978	48.4	592	2	Q2MLH9 HUMAN
14	874	43.3	597	2	Q4TB80 _TETNG
15	742.5	36.8	255	2	Q4TC88 _TETNG
16	622.5	30.8	515	2	Q8TI08 _BOMMO
17	407	20.1	543	1	HPSE1 ARATH
18	379.5	18.8	526	2	Q5SNA6 ORYSA
19	354.5	17.5	536	1	HPSE3 ARATH
20	354	17.5	401	2	Q30324 ARATH
21	352	17.4	559	2	Q89F99 BRAJZA
22	349.5	17.3	541	2	Q69116 ORYSA
23	349	17.3	539	2	Q2QN56 ORYSA
24	347.5	17.2	537	2	Q70YJ3 HORVU
25	345.5	17.1	539	1	HPSE2 ARATH
26	345	17.1	529	2	Q6ZJE2 ORYSA
27	339	16.8	527	2	Q9LRC8 SCUBA
28	290	14.4	516	2	Q447R5 SOLUS
29	237.5	11.8	506	2	Q37Q70 SPHAR
30	237	11.7	382	2	Q3E8P7 ARATH
31	236.5	11.7	537	2	Q43S03 SOLUS

32	156	7.7	935	2	Q9VE79 DROME
33	135	6.7	1128	2	Q5TT65 ANOGA
34	130	6.4	463	2	Q3IT97 BURPS
35	130	6.4	670	2	Q3JTG0 BURP1
36	126.5	6.3	493	2	Q9HK01 THEAC
37	124	6.1	510	2	Q2UOT3 ASPOR
38	123	6.1	634	2	Q5NJT7 _9BACT
39	122	6.0	559	2	Q7SPB0 NEUCR
40	119	5.9	815	2	Q54GJ2 DICDI
41	116	5.7	390	2	Q8TPH7 METAC
42	112.5	5.6	489	2	Q5DYH2 VIBF1
43	111	5.5	536	2	Q2UDS9 ASPOR
44	110.5	5.5	653	2	Q3NZS7 _9GAMM
45	110	5.4	665	2	Q55IC3 _CRNE

#### ALIGNMENTS

RESULT 1  
HPSE HUMAN  
ID HPSE\_HUMAN STANDARD; PRT; 543 AA.  
AC Q9Y251; Q53GE5; Q9UL39;  
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
DT 01-NOV-1999, sequence version 1.  
DT 07-FEB-2006, entry version 27.  
DE Heparanase precursor (EC 3.2.-.-) (Heparanase-1) (Hpa1) (Endo-glucuronidase) [Contains: Heparanase 8 kDa subunit; Heparanase 50 kDa subunit].  
GN Name=HPSE; Synonyms=HEP, HPA, HPAL, HPRI, HPSE1, HSE1;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.  
RC TISSUE=Placenta;  
RX MEDLINE=99335379; PubMed=10405343; DOI=10.1006/bbrc.1999.0962;  
RA Kussie P.H., Hulmes J.D., Ludwig D.L., Patel S., Navarro E.C., Seddon A.P., Giorgio N.A., Bohlen P.;  
RT "Cloning and functional expression of a human heparanase gene.";  
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).  
[2]  
RN NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICO-CHEMICAL PROPERTIES, AND PROTEIN SEQUENCE OF 158-168; 326-337 AND 447-491.  
RP SEQUENCE OF 158-168; 326-337 AND 447-491.  
RC TISSUE=Embryonic fibroblast;  
RX MEDLINE=99377052; PubMed=10446189; DOI=10.1074/jbc.274.34.24153;  
RA Toyoshima M., Nakajima M.;  
RT "Human heparanase. Purification, characterization, cloning, and expression.";  
RL J. Biol. Chem. 274:24153-24160(1999).  
[3]  
RN NUCLEOTIDE SEQUENCE [MRNA], N-GLYCOSYLATION, AND PROCESSING.  
RX PubMed=10395325; DOI=10.1038/10518;  
RA Vlodaysky I., Friedmann Y., Elkin M., Aingorn H., Atzmon R., Iehai-Michaeli R., Bitan M., Pappo O., Peretz T., Michael I., Spector L., Becker I.;  
RT "Mammalian heparanase: gene cloning, expression and function in tumor progression and metastasis.";  
RL Nat. Med. 5:793-802(1999).  
[4]  
RN NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND PROTEIN SEQUENCE OF 158-174; 263-272; 326-337; 433-436; 438-443; 466-468 AND 478-483.  
RC TISSUE=Placenta;  
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;  
RA Hullett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J., Parish C.R.;  
RT "Cloning of mammalian heparanase, an important enzyme in tumor invasion and metastasis.";  
RL Nat. Med. 5:803-809(1999).  
[5]  
RN NUCLEOTIDE SEQUENCE [MRNA], BIOPHYSICO-CHEMICAL PROPERTIES, AND TISSUE

RP TISSUE=Placenta;  
RX MEDLINE=20229546; PubMed=10764835; DOI=10.1093/glycob/10.5.467;  
RA Dempsey L.A., Plummer T.B., Coombes S.L., Platt J.L.;  
RT "Heparanase expression in invasive trophoblasts and acute vascular  
RT damage.";  
RL Glycobiology. 10:467-475(2000).  
RN [6]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND SUBCELLULAR LOCATION.  
RX PubMed=11547900; DOI=10.1023/A:1011375624902;  
RA Zcharia E., Metzger S., Chajek-Shaul T., Friedmann Y., Pappo O.,  
RA Aviv A., Elkin M., Pecker I., Peretz T., Vlodavsky I.;  
RT "Molecular properties and involvement of heparanase in cancer  
RT progression and mammary gland morphogenesis.";  
RL J. Mammary Gland Biol. Neoplasia 6:311-322(2001).  
RN [7]  
RP NUCLEOTIDE SEQUENCE [MRNA], PROTEIN SEQUENCE OF 36-41 AND 158-163,  
RP SUBUNITS, GLYCOSYLATION, AND BIOPHYSICOCHEMICAL PROPERTIES.  
RX TISSUE=Placenta; DOI=10.1042/BJ20030318;  
RA McKenzie E., Young K., Hircok M., Bennett J., Bhaman M., Felix R.,  
RA Turner P., Stamps A., McMillan D., Saville G., Ng S., Mason S.,  
RA Snell D., Schofield D., Gong H., Townsend R., Gallagher J., Page M.,  
RA Parakh R., Stubberfield C.;  
RT "Biochemical characterization of the active heterodimer form of human  
RT heparanase (Hpal) protein expressed in insect cells.";  
RL Biochem. J. 373:423-435(2003).  
RN [8]  
RP NUCLEOTIDE SEQUENCE [MRNA].  
RA Pinhal M.A., Semedo P.;  
RT "Cloned heparanase from MCF-7 cells.";  
RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.  
RN [9]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RX TISSUE=Small intestine;  
RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,  
RA Tanaka A., Yokoyama S.;  
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.  
RN [10]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RX TISSUE=Pancreas;  
RA MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Atschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [11]  
RP MUTAGENESIS OF GLU-225; GLU-343 AND ASP-367.  
RX PubMed=11123890; DOI=10.1021/b1002080p;  
RA Hulett M.D., Hornby J.R., Ohms S.J., Zuegg J., Freeman C.,  
RA Gready J.E., Parish C.R.;  
RT "Identification of active-site residues of the pro-metastatic  
RT endoglycosidase heparanase.";  
RL Biochemistry 39:15659-15667(2000).  
RN [12]  
RP N-GLYCOSYLATION, AND MUTAGENESIS OF ASN-162; ASN-178; ASN-200;  
RP ASN-217; ASN-238 AND ASN-459.

RX PubMed=14573609; DOI=10.1074/jbc.M300541200;  
RA Simizu S., Ishida K., Wierzb M.K., Osada H.;  
RT "Secretion of heparanase protein is regulated by glycosylation in  
RT human tumor cell lines.";  
RL J. Biol. Chem. 279:2697-2703(2004).  
RN [13]  
RP SUBCELLULAR LOCATION.  
RX PubMed=15292202; DOI=10.1074/jbc.M402131200;  
RA Gangis-Velitski S., Zetser A., Kaplan V., Ben-zaken O., Cohen E.,  
RA Levy-Adam F., Bashenko Y., Flugelman M.Y., Vlodavsky I., Ilan N.;  
RT "Heparanase uptake is mediated by cell membrane heparan sulfate  
RT proteoglycans.";  
RL J. Biol. Chem. 279:44084-44092(2004).  
RN [14]  
RP BIOPHYSICOCHEMICAL PROPERTIES, PROCESSING, AND SUBCELLULAR LOCATION.  
RX PubMed=15848168; DOI=10.1016/j.febslet.2005.03.030;  
RA Cohen E., Atzmon R., Vlodavsky I., Ilan N.;  
RT "Heparanase processing by lysosomal/endosomal protein preparation.";  
RL FEBS Lett. 579:2334-2338(2005).  
RN [15]  
RP SUBCELLULAR LOCATION, PROCESSING, AND MUTAGENESIS OF TYR-156.  
RX PubMed=15659389; DOI=10.1074/jbc.M413370200;  
RA Abboud-Jarrous G., Rangini-Guetta Z., Aingorn H., Atzmon R.,  
RA Elgavish S., Peretz T., Vlodavsky I.;  
RT "Site-directed mutagenesis, proteolytic cleavage, and activation of  
RT human proheparanase.";  
RL J. Biol. Chem. 280:13568-13575(2005).  
RN [16]  
RP DOMAINS, AND MUTAGENESIS OF LYS-158 AND LYS-161.  
RX PubMed=15760902; DOI=10.1074/jbc.M414546200;  
RA Levy-Adam F., Abboud-Jarrous G., Guerrini M., Beccati D.,  
RA Vlodavsky I., Ilan N.;  
RT "Identification and characterization of heparin/heparan sulfate  
RT binding domains of the endoglycosidase heparanase.";  
RL J. Biol. Chem. 280:20457-20466(2005).  
RN [17]  
RP VARIANT SER-260.  
RX PubMed=15334672;  
RA Chen X.P., Liu Y.B., Rui J., Peng S.Y., Peng C.H., Zhou Z.Y.,  
RA Shi L.H., Shen H.W., Xu B.;  
RT "Heparanase mRNA expression and point mutation in hepatocellular  
RT carcinoma.";  
RL World J. Gastroenterol. 10:2795-2799(2004).  
RN [18]  
CC -I- FUNCTION: Endoglycosidase which is a cell surface and  
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
CC proteoglycans. Also implicated in the extravasation of leukocytes  
CC and tumor cell lines. Due to its contribution to metastasis and  
CC angiogenesis, it is considered to be a potential target for anti-  
CC cancer therapies.  
CC -I- ENZYME REGULATION: Inhibited by EDTA, laminarin sulfate and, to a  
CC lower extent, by heparin and sulfamin and activated by calcium and  
CC magnesium (By similarity).  
CC -I- BIOPHYSICOCHEMICAL PROPERTIES:  
CC pH dependence:  
CC Optimum pH is 4-6;  
CC -I- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
CC subunits, the proteolytic products.  
CC -I- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.  
CC Secreted, internalised and transferred to late endosomes/lysosomes  
CC as a proheparanase. In lysosomes, it is processed into the active  
CC form, the heparanase. The uptake or internalisation of  
CC proheparanase is mediated by HSPGs. Heparin appears to be a  
CC competitor and retain proheparanase in the extracellular medium.  
CC -I- TISSUE SPECIFICITY: Highly expressed in placenta and spleen and  
CC weakly expressed in lymph node, thymus, peripheral blood  
CC leukocytes, bone marrow, endothelial cells, fetal liver and tumor  
CC tissues.  
CC -I- PTM: Proteolytically processed. The cleavage of the 65 kDa form  
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa  
CC product. The active form, the 8/50 kDa heterodimer, is resistant  
CC to degradation. Complete removal of the linker peptide appears to  
CC be a prerequisite to the complete activation of the enzyme.

```
CC -I- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC to be essential for its solubility.

Query Match      100.0%; Score 2020; DB 1; Length 543;
Best Local Similarity 100.0%; Pred. No. 3.3e-154;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLOWNSSNAQLLDYCSSKGYN 60
DB 158 KKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLOWNSSNAQLLDYCSSKGYN 217
QY 61 ISWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 120
DB 218 ISWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 277
QY 121 MLKSFLLKAGEVIDSVTWHYYLNGRTATREDFLNPDVLDIFISSVQVQFVVESTRPGK 180
DB 278 MLKSFLLKAGEVIDSVTWHYYLNGRTATREDFLNPDVLDIFISSVQVQFVVESTRPGK 337
QY 181 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMQVFFGAGNYHLVDEN 240
DB 338 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMQVFFGAGNYHLVDEN 397
QY 241 FDPDPYWLSSLFVKLVGTGKVLMAVSGSKRRKLRVYLHCTNTDNPYKEGDITLYAINL 300
DB 398 FDPDPYWLSSLFVKLVGTGKVLMAVSGSKRRKLRVYLHCTNTDNPYKEGDITLYAINL 457
QY 301 HNVTKYLRPLPYPSNKKQVDKYLRLPLGPHGLLSKSVQLNGLTLKMWDDQTLPLMEKPLR 360
DB 458 HNVTKYLRPLPYPSNKKQVDKYLRLPLGPHGLLSKSVQLNGLTLKMWDDQTLPLMEKPLR 517
QY 361 PGSSGLGPAFSYFFVIRNAKVAACI 386
DB 518 PGSSGLGPAFSYFFVIRNAKVAACI 543

RESULT 2
Q333X9_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X9;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax galili.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164323;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
RN [2];
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; AM085490; CAJ30017.1; -; mRNA.
SQ SEQUENCE 574 AA; 64525 MW; 1635865051B380D0 CRC64;

Query Match      85.3%; Score 1724; DB 2; Length 574;
Best Local Similarity 84.5%; Pred. No. 2.9e-130;
Matches 325; Conservative 30; Mismatches 31; Indels 0; Gaps 0;

QY 1 KKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLOWNSSNAQLLDYCSSKGYN 60
DB 189 KKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLOWNSSNAQLLDYCSSKGYN 248
```

```
Matches 326; Conservative 28; Mismatches 32; Indels 0; Gaps 0;

QY 1 KKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLOWNSSNAQLLDYCSSKGYN 60
DB 189 KKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLOWNSSNAQLLDYCSSKGYN 248
QY 61 ISWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 120
DB 249 ISWELGNEPNSFLKKADIFINGSQLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 308
QY 121 MLKSFLLKAGEVIDSVTWHYYLNGRTATREDFLNPDVLDIFISSVQVQFVVESTRPGK 180
DB 309 LRSFLKAGGEVIDSVTWHYYLNGRIATKEDFLSPVDLDTFILSVQKILQVVESTRPGK 368
QY 181 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMQVFFGAGNYHLVDEN 240
DB 369 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMQVFFGAGNYHLVDEN 428
QY 241 FDPDPYWLSSLFVKLVGTGKVLMAVSGSKRRKLRVYLHCTNTDNPYKEGDITLYAINL 300
DB 429 FEPLPDYWLSSLFVKLVGTGKVLMAVSGSKRRKLRVYLHCTNTDNPYKEGDITLYAINL 488
QY 301 HNVTKYLRPLPYPSNKKQVDKYLRLPLGPHGLLSKSVQLNGLTLKMWDDQTLPLMEKPLR 360
DB 489 YNVTGHLKLPYQLFNKPVDKYLVKPLGPGGLLSKSVQLNGLTLKMWDDQTLPLMEKPLR 548
QY 361 PGSSGLGPAFSYFFVIRNAKVAACI 386
DB 549 PGSSGLGPAFSYFFVIRNAKVAACI 574

RESULT 3
Q333X8_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X8;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Heparanase.
GN Name=hpa;
OS Spalax golani.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=191382;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
RN [2];
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avivi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; AM085491; CAJ30018.1; -; mRNA.
SQ SEQUENCE 574 AA; 64555 MW; 48EBFEFC7D0BCB34 CRC64;

Query Match      85.3%; Score 1723; DB 2; Length 574;
Best Local Similarity 84.2%; Pred. No. 3.5e-130;
Matches 325; Conservative 30; Mismatches 31; Indels 0; Gaps 0;

QY 1 KKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLOWNSSNAQLLDYCSSKGYN 60
DB 189 KKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLOWNSSNAQLLDYCSSKGYN 248
```



```
QY 61 ISWELGNPNPSFLKADIPIFNGSOLGEDFTQLHLKLRKSTFKNAKLYGPDVGQPRKTKA 120
DB 249 ISWELGNPNPSFWKKAHISIDGLQGEDYIELKLRKSTLKNVLYGPDVGQPRGKTVK 308
QY 121 MLKSLFKAGGEVIDSVTHHHYILNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGK 180
DB 309 LRSFLKAGGEVIDSVTHHHYILNGRIATKEDFLSPDLDTFLSVQKILQVVESTRPGK 368
QY 181 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
DB 369 KWLGETSSAYGGAPLLSNTFAAGFMWLDKLGLSQMGIEVVMRQVFFGAGNYHLVDKN 428
QY 241 FDPDPYWLSSLFKLVGTVKVLMAVSGSKRRKLRVYLHCTNTDNPRYKEGDGLTYALNL 300
DB 429 FEPLDPYWLSSLFKLVGSKVLMARVKGPDORSKLRVYLHCTNINHPRYQEGDGLTYALNL 488
QY 301 HNVTKYLRPLYPSPNKQVDKYLRLPLGPHGLSKSVQNLGLTLKMVDOTLPLMEKPLR 360
DB 489 YNVTXHLKLPYQLENKPDVKYLKPLGPGGLSKSVQNLQALKMVDOTLPLATEKPLR 548
QY 361 PGSSSLGLPAFYSYFFVIRNAKVAACI 386
DB 549 PGSSSLGLPAFYSYFFVIRNAKVAACL 574

RESULT 4
Q333X7_9RODE PRELIMINARY; PRT; 574 AA.
AC Q333X7_9RODE
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, sequence version 1.
DE Heparanase.
GN Name=hpa;
OS Spalax carmeli.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=164324;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DR EMBL; AM085492; CAJ30019.1; -; mRNA.
SQ SEQUENCE 574 AA; 64459 MW; 9FID19DCBADD99DE CRC64;

Query Match 84.7%; Score 1711; DB 2; Length 574;
Best Local Similarity 83.9%; Pred. No. 3.3e-129;
Matches 324; Conservative 29; Mismatches 33; Indels 0; Gaps 0;

QY 1 KKFKNSTYSSRSVDVLYTFANGSGLDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYN 60
DB 189 KEFKNSTYSSRSVDVLYTFARCSGLDLIFGLNALLRTADFRWNSNAQLLNYCSSKNYD 248
QY 61 ISWELGNPNPSFLKADIPIFNGSOLGEDFTQLHLKLRKSTFKNAKLYGPDVGQPRKTKA 120
DB 249 ISWELGNPNPSFWKKAHISIDGLQGEDYIELKLRKSTLKNVLYGPDVGQPRGKTVK 308
QY 121 MLKSLFKAGGEVIDSVTHHHYILNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGK 180
DB 309 LRSFLKAGGEVIDSVTHHHYILNGRIATKEDFLSPDLDTFLSVQKILQVVESTRPGK 368
QY 181 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
DB 369 KWLGETSSAYGGAPLLSNTFAAGFMWLDKLGLSQMGIEVVMRQVFFGAGNYHLVDKN 428
QY 241 FDPDPYWLSSLFKLVGTVKVLMAVSGSKRRKLRVYLHCTNTDNPRYKEGDGLTYALNL 300
DB 429 FEPLDPYWLSSLFKLVGSKVLMARVKGPDORSKLRVYLHCTNINHPRYQEGDGLTYALNL 488
```

```
QY 241 FDPDPYWLSSLFKLVGTVKVLMAVSGSKRRKLRVYLHCTNTDNPRYKEGDGLTYALNL 300
DB 429 FEPLDPYWLSSLFKLVGSKVLMARVKGPDORSKLRVYLHCTNINHPRYQEGDGLTYALNL 488
QY 301 HNVTKYLRPLYPSPNKQVDKYLRLPLGPHGLSKSVQNLGLTLKMVDOTLPLMEKPLR 360
DB 489 YNVTXHLKLPYQLENKPDVKYLKPLGPGGLSKSVQNLQALKMVDOTLPLATEKPLR 548
QY 361 PGSSSLGLPAFYSYFFVIRNAKVAACI 386
DB 549 PGSSSLGLPAFYSYFFVIRNAKVAACL 574

RESULT 5
Q333X6_SPAJD PRELIMINARY; PRT; 574 AA.
AC Q333X6_SPAJD
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, sequence version 1.
DE Heparanase.
GN Name=hpa;
OS Spalax judaei (Blind subterranean mole rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Spalacidae; Spalacinae; Spalax.
OX NCBI_TaxID=134510;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avi A.;
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene
cloning and identification of a novel splice variant.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DR EMBL; AM085493; CAJ30020.1; -; mRNA.
SQ SEQUENCE 574 AA; 64515 MW; 3AEBB13F07451684 CRC64;

Query Match 84.7%; Score 1710; DB 2; Length 574;
Best Local Similarity 83.9%; Pred. No. 3.9e-129;
Matches 324; Conservative 29; Mismatches 33; Indels 0; Gaps 0;

QY 1 KKFKNSTYSSRSVDVLYTFANGSGLDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYN 60
DB 189 KEFKNSTYSSRSVDVLYTFARCSGLDLIFGLNALLRTADFRWNSNAQLLNYCSSKNYD 248
QY 61 ISWELGNPNPSFLKADIPIFNGSOLGEDFTQLHLKLRKSTFKNAKLYGPDVGQPRKTKA 120
DB 249 ISWELGNPNPSFWKKAHISIDGLQGEDYIELKLRKSTLKNVLYGPDVGQPRGKTVK 308
QY 121 MLKSLFKAGGEVIDSVTHHHYILNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGK 180
DB 309 LRSFLKAGGEVIDSVTHHHYILNGRIATKEDFLSPDLDTFLSVQKILQVVESTRPGK 368
QY 181 KWLGETSSAYGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
DB 369 KWLGETSSAYGGAPLLSNTFAAGFMWLDKLGLSQMGIEVVMRQVFFGAGNYHLVDKN 428
QY 241 FDPDPYWLSSLFKLVGTVKVLMAVSGSKRRKLRVYLHCTNTDNPRYKEGDGLTYALNL 300
DB 429 FEPLDPYWLSSLFKLVGSKVLMARVKGPDORSKLRVYLHCTNINHPRYQEGDGLTYALNL 488
```



Qy	301	HNVTKYLRPLPFPFSKNQVDKYLRLRPLGPGHLLSKSVQLNGLTLKWVDDQTLPALTEKPUR	360
Dd	489	YNVTGHKLPLPYQLFNPKVPDKYLVIPLPGGLLSKSVQLNGQALKWKVDDQTLPALTEKPUR	548
Qy	361	PGSSILGLPAFYSYFFVIIRNAKVAACI	386
Dd	549	PGSSILGLPAFYSYGFFVIIRNAKVAACL	574
RESULT 6			
ID	HPSE_BOVIN	STANDARD;	PRT; 545 AA.
AC	Q9MYO,		
DT	11-OCT-2005,	integrated into UniProtKB/Swiss-Prot.	
DT	01-JUN-2001,	sequence version 2.	
DT	07-MAR-2006,	entry version 15.	
DE	Heparanase precursor (EC 3.2.-.-)	[Contains: Heparanase 8 kDa subunit;	
DE	Heparanase	50 kDa subunit].	
GN	Name=HPSE;		
OS	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;		
OC	Pecora; Bovidae; Bovinae; Bos.		
OX	NCBI_TaxID=9913;		
CC	[1]		
RN	NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.		
RP	TISSUE=Placenta;		
RC	MEDLINE=21176669; PubMed=112776877;		
RA	Kizaki K., Nakano H., Nakano H., Takahashi T., Imai K., Hashizume K.;		
RX	"Expression of heparanase mRNA in bovine placenta during gestation.";		
RT	Reproduction 121:573-580(2001).		
RL	- - FUNCTION: Endoglycosidase which is a cell surface and extracellular matrix-degrading enzyme. Cleaves heparan sulfate proteoglycans (HSPGs) into heparan sulfate side chains and core proteoglycans. Also implicated in the extravasation of leukocytes and tumor cell lines. Contributes to metastasis and angiogenesis (By similarity).		
CC	- - ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower extent, by heparin, sulfamin and EDTA. Activated by calcium and magnesium (By similarity).		
CC	- - SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa subunits, the proteolytic products (By similarity).		
CC	- - SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted. Secreted, internalised and transferred to late endosomes/lysosomes as a proheparanase. In lysosomes, it is processed into the active form, the heparanase. The uptake or internalisation of proheparanase is mediated by HSPGs. Heparin appears to be a competitor and retain proheparanase in the extracellular medium (By similarity).		
CC	- - TISSUE SPECIFICITY: Highly expressed in placenta and weakly in the kidney, lung, spleen and uterus.		
CC	- - PM: Proteolytically processed. The cleavage of the 65 kDa form leads to the generation of a linker peptide, 8 kDa and 50 kDa product. The active form, the 8/50 kDa heterodimer, is resistant to degradation. Complete removal of the linker peptide appears to be a prerequisite to the complete activation of the enzyme (By similarity).		
CC	- - PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears to be essential for its solubility (By similarity).		
CC	- - SIMILARITY: Belongs to the glycosyl hydrolase 79 family.		
-----			
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms Distributed under the Creative Commons Attribution-NoDerivs License		
-----			
DR	EMBL; AF281160; AAF87301.2; -; mRNA.		
DR	InterPro; IPR005199; Glyco_hydro_79_N.		
DR	Pfam; PF03662; Glyco_hydro_79n; 1.		
DR	CaSc; Glycoprotein; Hydrolase; Lysozyme; Magnesium; Membrane; Signal.		
KW	SIGNAL		
FT	CHAIN	1 37 By similarity.	
FT	CHAIN	38 111 Heparanase 8 kDa subunit (By similarity).	
FT	PROPEP	112 159 Linker peptide.	
FT			

RP STRAIN=FBV; TISSUE=Embryo;  
RX MEDLINE=22350326; PubMed=12460766; DOI=10.1016/S1046-5928(02)00558-2;  
RA Miao H.-Q., Navarro E., Patel S., Sargent D., Koo H., Wan H.,  
RA Plata A., Zhou Q., Ludwig D., Bohlen P., Kussie P.;  
RT "Cloning, expression, and purification of mouse heparanase.";  
RL Protein Expr. Purif. 26:425-431(2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.  
RX MEDLINE=22841152; PubMed=12837765; DOI=10.1074/jbc.M300925200;  
RA Gong F., Jemth P., Galvis M.L.E., Vlodavsky I., Horner A., Lindahl U.,  
RA Li J.-P.;  
RT "Processing of macromolecular heparin by heparanase.";  
RL J. Biol. Chem. 278:35152-35158(2003).  
RN [4]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC STRAIN=C57BL/6J, and NOD; TISSUE=Thymus;  
RX PubMed=16141072; DOI=10.1126/science.1111014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
RA Ambesi-Impombato A., Aweller R., Aturaliya R.N., Bailey T.L.,  
RA Banal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Humnick L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Mottaqui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J.F., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,  
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Inamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashima T., Kojima M., Kondou S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Pleassy C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563(2005).  
CC -!- FUNCTION: Endoglycosidase which is a cell surface and  
extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
proteoglycans (HSPGs) into heparan sulfate side chains and core  
proteoglycans. Also implicated in the extravasation of leukocytes  
and tumor cell lines. Contributes to metastasis and angiogenesis  
(By similarity).  
CC -!- ENZYME REGULATION: Inhibited by EDTA and activated by calcium and  
magnesium (By similarity). Inhibited by laminarin sulfate and, to  
a lower extent, by heparin and sulfamin.  
CC -!- BIOPHYSICOCHEMICAL PROPERTIES:  
pH dependence:  
Optimum pH is 5;  
CC -!- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
subunits, the proteolytic products.  
CC -!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.  
CC Secreted, internalised and transferred to late endosomes/lysosomes

CC	as a proheparanase. In lysosomes, it is processed into the active
CC	form, the heparanase. The uptake or internalisation of
CC	proheparanase is mediated by HSPGs. Heparin appears to be a
CC	competitor and retain proheparanase in the extracellular medium
CC	(By similarity).
CC	-!- PTM: Proteolytically processed. The cleavage of the 65 kDa form
CC	leads to the generation of a linker peptide, 8 kDa and 50 kDa
CC	product. The active form, the 8/50 kDa heterodimer, is resistant
CC	to degradation. Complete removal of the linker peptide appears to
CC	be a prerequisite to the complete activation of the enzyme (By
CC	similarity).
CC	-!- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears
CC	to be essential for its solubility.
CC	-!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC	Copyrighted by the Uniprot Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a>
CC	Distributed under the Creative Commons Attribution-NoDerivs License
CC	-----
DR	EMBL; AF395907; AAQ15188.1; -; mRNA.
DR	EMBL; AY077467; AAL76083.1; -; mRNA.
DR	EMBL; AX151051; AAN41636.1; -; mRNA.
DR	EMBL; AK040471; BAC30600.1; -; mRNA.
DR	EMBL; AK154628; BAE32725.1; -; mRNA.
DR	Ensembl; ENSMUSG00000035273; Mus musculus.
DR	MGI; MGI:1343124; Hpsc.
DR	GO; GO:0005578; C:extracellular matrix (sensu Metazoa); TAS.
DR	InterPro; IPR005199; Glyco_hydro_79_N.
DR	Pfam; PF03662; Glyco_hydro_79n; 1.
DR	Calcium; Direct protein sequencing; Glycoprotein; Hydrolase; Lysosome;
KW	Magnesium; Membrane; Signal.
FT	CHAIN 1 27
FT	By similarity.
FT	PROPEP 102 149
FT	By similarity.
FT	CHAIN 150 535
FT	Heparanase 8 kDa subunit.
FT	/FTid=PRO_0000042263.
FT	Linker peptide (By similarity).
FT	/FTid=PRO_0000042264.
FT	Heparanase 50 kDa subunit.
FT	/FTid=PRO_0000042265.
FT	Heparin/HS-binding (By similarity).
FT	REGION 150 154
FT	Heparin/HS-binding (By similarity).
FT	ACT_SITE 262 272
FT	Proton donor (Potential).
FT	ACT_SITE 335 335
FT	Nucleophile (Potential).
FT	CARBOHYD 154 154
FT	N-linked (GlcNAc. . .)
FT	CARBOHYD 192 192
FT	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD 209 209
FT	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD 230 230
FT	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD 451 451
FT	K -> R (in Ref. 3).
FT	CONFLICT 206 206
FT	W -> S (in Ref. 3).
FT	CONFLICT 212 212
FT	NGS -> DGL (in Ref. 1, 2 and 4).
FT	CONFLICT 230 232
FT	E -> K (in Ref. 3).
FT	CONFLICT 335 335
FT	G -> A (in Ref. 3).
FT	CONFLICT 342 342
FT	Y -> H (in Ref. 1, 2 and 4).
FT	CONFLICT 455 455
FT	V -> I (in Ref. 1, 2 and 4).
FT	CONFLICT 531 531
FT	AF19E287BCD03F7B CRC64;
SQ	SEQUENCE 535 AA; 60050 MW; AF19E287BCD03F7B CRC64;
Query Match	82.2%; Score 1661; DB 1; Length 535;
Best Local Similarity	80.8%; Pred. No. 3,2e-125;
Matches 312; Conservative	33; Mismatches 41; Indels 0; Gaps 0;
QY	1 KKFKNSTYSRSDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYN 60
Db	150 KEFKNSTYSRSDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYN 209
QY	61 ISWELGNPNFLKADIFINGSLQGLDFQLHKLKSTFKNAKLYGPDVGQPRRTAK 120
Db	210 ISWELGNPNFLKADIFINGSLQGLDFQLHKLKSTFKNAKLYGPDVGQPRRTAK 269
QY	121 MLKSLFKAGGEVIDSVTHHYLYNGRTATREDFLNPDVLDFISVSVKQVFOVWSETRPGK 180
Db	270 LKSLFKAGGEVIDSVTHHYLYNGRTATREDFLNPDVLDFISVSVKQVFOVWSETRPGK 329
QY	181 KWLIGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 240

Db 330 KWLGETSSAYGGGAPLLSNTFAAGFMWLDKLGLSAQMGIEVVMRQVFFGAGNHYLVN 389  
Qy 241 FDPDPYWLSSLFVKLVGTVMASVGGSKRKLRVYLHCTNTDNPYKGGDLTYAINL 300  
Db 390 FEPLDPYWLSSLFVKLVGPRVLSRVKGPDRSKLRVYLHCTNVVHPRYQGGDLTYLVNL 449  
Qy 301 HNVTKYLRPLPYPSNQVDKYLRLPGLPHGLLSKSVQLNGLTLMKVVDDQTLPLMEKPLR 360  
Db 450 HNVTKYLVKPPPLFRKFPVDYLLKPSGPDGLLSKSVQLNGLTLMKVVDDQTLPLMEKPLR 509  
Qy 361 PGSSGLGPAFSPSYFFVIRNAKVAACI 386  
Db 510 AGSALSPLAFSYGFFVIRNAKVAACI 535

RESULT 8  
HPSE RAT  
ID HPSE RAT STANDARD; PRT; 536 AA.  
AC Q71RPL; Q90ZF8;  
DT 11-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
DT 05-JUL-2004, sequence version 1.  
DT 07-MAR-2006, entry version 1.  
DE Heparanase precursor (EC 3.2.-.-) (Endo-glucuronidase) (Contains:  
DE Heparanase 8 kDa subunit; Heparanase 50 kDa subunit).  
GN Name=Hps; Synonyms=Hep;  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN (1)\_TaxID=10116;  
RP NUCLEOTIDE SEQUENCE [MRNA].  
RC TISSUE=Placenta;  
RX MEDLINE=99321249; PubMed=10395326; DOI=10.1038/10525;  
RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,  
RA Parish C.R.;  
RT "Cloning of mammalian heparanase, an important enzyme in tumor  
RT invasion and metastasis."  
RL Nat. Med. 5:803-809(1999).  
RN (2)  
RP NUCLEOTIDE SEQUENCE [MRNA], AND ENZYME REGULATION.  
RX MEDLINE=22194309; PubMed=12077130; DOI=10.1074/jbc.M203282200;  
RA Podyna-Inoue K.A., Yokote H., Sakaguchi K., Ikuta M., Yangishita M.;  
RT "Characterization of heparanase from a rat parathyroid cell line."  
RL J. Biol. Chem. 277:32459-32465(2002).  
CC !- FUNCTION: Endoglycosidase which is a cell surface and  
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate  
CC proteoglycans (HSPGs) into heparan sulfate side chains and core  
CC proteoglycans. Also implicated in the extravasation of leukocytes  
CC and tumor cell lines. Contributes to metastasis and angiogenesis  
CC (By similarity).  
CC !- ENZYME REGULATION: Inhibited by laminarin sulfate and, to a lower  
CC extent, by heparin and sulfamin (By similarity). Activated by  
CC calcium and magnesium. Inhibited by EDTA.  
CC !- SUBUNIT: The active heterodimer is composed of the 8 and 50 kDa  
CC subunits, the proteolytic products (By similarity).  
CC !- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted.  
CC Secreted, internalised and transferred to late endosomes/lysosomes  
CC as a proheparanase. In lysosomes, it is processed into the active  
CC form, the heparanase. The uptake or internalisation of  
CC proheparanase is mediated by HSPGs. Heparin appears to be a  
CC competitor and retain proheparanase in the extracellular medium  
CC (By similarity).  
CC !- PTM: Proteolytically processed. The cleavage of the 65 kDa form  
CC leads to the generation of a linker peptide, 8 kDa and 50 kDa  
CC product. The active form, the 8/50 kDa heterodimer, is resistant  
CC to degradation. Complete removal of the linker peptide appears to  
CC be a prerequisite to the complete activation of the enzyme (By  
CC similarity).  
CC !- PTM: N-glycosylated. Glycosylation of the 50 kDa subunit appears  
CC to be essential for its solubility (By similarity).  
CC !- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AF59508; AAQ15189.1; -, mRNA.  
DR EMBL; AF184967; AAF04563.1; -, mRNA.  
DR RGD; 61969; Hps.  
DR InterPro; IPR005199; Glyco\_hydro\_79\_N.  
DR Pfam; PF03662; Glyco\_hydro\_79n; 1.  
KW Calcium; Glycoprotein; Hydrolase; Lysosome; Magnesium; Membrane;  
KW Signal.  
FT SIGNAL 1 28 By similarity.  
FT CHAIN 29 102 Heparanase 8 kDa subunit.  
FT PROPEP 103 150 Linker peptide (By similarity).  
FT CHAIN 151 536 Heparanase 50 kDa subunit.  
FT REGION 151 155 Heparin/HS-binding (By similarity).  
FT REGION 263 273 Heparin/HS-binding (By similarity).  
FT ACT\_SITE 218 218 Proton donor (Potential).  
FT ACT\_SITE 336 336 Nucleophile (Potential).  
FT CARBOHYD 155 155 N-linked (GlcNAc...) (By similarity).  
FT CARBOHYD 193 193 N-linked (GlcNAc...) (By similarity).  
FT CARBOHYD 210 210 N-linked (GlcNAc...) (By similarity).  
FT CARBOHYD 452 452 N-linked (GlcNAc...) (By similarity).  
FT CONFLICT 15 15 G -> R (in Ref. 2).  
FT CONFLICT 227 227 H -> Q (in Ref. 2).  
FT CONFLICT 350 350 D -> N (in Ref. 2).  
SQ SEQUENCE 536 AA; 60480 MW; C43E04CF536EA4D CRC64;  
Query Match 81.5%; Score 1646; DB 1; Length 536;  
Best Local Similarity 80.1%; Pred. No. 5.3e-124;  
Matches 309; Conservative 36; Mismatches 41; Indels 0; Gaps 0;  
Qy 1 KFKNNTSYSSVDVLTFTACSGLDLIFGLNALLRTADLQWSSNAQLLDYCSKGYN 60  
Db 151 REFKNSTYSRSSVDMLYSFAKCSRLDLIFGLNALLRTPDLRWSSNAQLLDYCSKGYN 210  
Qy 61 ISWELGNPNPSFLKKADIFINGSQGLDFEVLHKLKRSFTFNKALYGPDVQPPRRKTA 120  
Db 211 ISWELGNPNPSFWKKAHISIDGLQGLDFEVLHKLKRSFAQNAKLYGPDIGQPPRKTVK 270  
Qy 121 MLKSLFKAGGEVDSTVTHHYLYNGRATREDFLNPVDLDFISVQKVFQVESTTRGK 180  
Db 271 LLRSFLKAGGEVDSTVTHHYLYNGRVATKEDFLSSVDLDTFLLSVQKILKVTKEMTGK 330  
Qy 181 KVMLETSSAYGGGAPLLSDTFAAGFMWLDKLGLSAQMGIEVVMRQVFFGAGNHYLVN 240  
Db 331 KVMLETSSAYGGGAPLLSDTFAAGFMWLDKLGLSAQMGIEVVMRQVFFGAGNHYLVN 390  
Qy 241 FDPDPYWLSSLFVKLVGTVMASVGGSKRKLRVYLHCTNTDNPYKGGDLTYAINL 300  
Db 391 FEPLDPYWLSSLFVKLVGPRVLSRVKGPDRSKLRVYLHCTNVVHPRYQGGDLTYLVNL 450  
Qy 301 HNVTKYLRPLPYPSNQVDKYLRLPGLPHGLLSKSVQLNGLTLMKVVDDQTLPLMEKPLR 360  
Db 451 HNVTKYLVKPPPLFRKFPVDYLLKPSGPDGLLSKSVQLNGLTLMKVVDDQTLPLMEKPLR 510  
Qy 361 PGSSGLGPAFSPSYFFVIRNAKVAACI 386  
Db 511 AGSALSPLAFSYGFFVIRNAKVAACI 536  
RESULT 9  
Q333X5\_SPAJD  
ID Q333X5\_SPAJD PRELIMINARY; PRT; 558 AA.  
AC Q333X5;  
DT 06-DEC-2005, integrated into UniProtKB/TREMBL.  
DT 06-DEC-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Heparanase.  
GN Name=hpa;  
OS Spalax judaei (Blind subterranean mole rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Spalacidae; Spalacinae; Spalax.  
ON NCBI\_TaxID=134510;  
RX [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Kidney;  
RA Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
RT cloning and identification of a novel splice variant."; Proc. Natl. Acad. Sci. U.S.A. 102:15161-15166(2005).  
RN [2]  
RL NUCLEOTIDE SEQUENCE.  
RP TISSUE=Kidney;  
RC Nasser N.J., Nevo E., Shafat I., Ilan N., Vlodavsky I., Avi A.;  
RT "Adaptive evolution of heparanase in hypoxia-tolerant Spalax: gene  
RT cloning and identification of a novel splice variant."; Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).  
CC -----  
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR ENBL; AM085494; CAJ30021.1; -; mRNA.  
SQ SEQUENCE 558 AA; 62737 MW; 0VBAF8F55849EE7 CRC64;

Query Match 79.3%; Score 1602; DB 2; Length 558;  
Best Local Similarity 79.8%; Pred. No. 2e-120;  
Matches 308; Conservative 29; Mismatches 33; Indels 16; Gaps 1;

QY	1	KKPKNSTYSRSSVDVLYTFANCGLDILFGNALLRRTADLQWNSNAQLLLDYCSSGYN	60
DB	189	KEFKNSTYSRSSVDMLYTFARCSGLDLIFGLNALLRRTADFWRNSSAQLLLNYCSSKYD	248
QY	61	ISWELGNPNFSFLKKADIFINGSOLGEDFTQLHKLRKSTFKNAKLPGPDVGQPRRTAK	120
DB	249	ISWELGNPNFSFWKKAHISIDGLQGSDYTELRKLLKSTLKNNKLYGPDVGQPRGTVK	308
QY	121	MLSKFLKAGGEVIDSVTWHYYLNGRTRATREDFLNPVDLDIFITSSQKVFOVVESTP GK	180
DB	309	LLRS-----YYLNGRTRATEDFLSPVDLTFLVSQKILQVZETP GK	352
QY	181	KWLGETSSAYGGAPLLSDTTFAAGFMWKDKLGLSARMGIEVNVRQVFPGAGNYHLVDEN	240
DB	353	KWLGETSSAYGGAPLLSNTFAAGFMWKDKLGLSAQMGIEVNVRQVFPGAGNYHLVDKN	412
QY	241	FDPDPDWLSLLFKKLYGTVKMVASVOGSKRRKLRVYLHCTNTDNPRYKSGDLLTYAINL	300
DB	413	FEPLPDWLSLLFKCLYGSKVMARVKGPPRSKLRVYLHCTNINHPRYQEGDLLTYALNL	472
QY	301	HNYTKYLRLPYPSNKQVDKYLRLPLGP HGLLSKSVLQNGLTTLKWVDDQTLPPLMEKPLR	360
DB	473	YNTKHLLPYQLFNKPVDKYLVIPLGPGGLLSKSVQLNGQALKWVDDQTLPALTETPLR	532
QY	361	PGSSLGLPAFYSYGFVVIRNAKVAACI	386
DB	533	PGSSLGLPAFYSYGFVVIRNAKVAACL	558

RESULT 10  
HPSE\_CHK ID HPSE\_CHK STANDARD; PRT; 523 AA.

AC	O30VK5;
DT	11-OCT-2005, integrated into UniProtKB/Swiss-Prot.
DT	01-DEC-2001, sequence version 1.
DE	07-FEB-2006, entry version 12.
DE	Heparanase precursor (EC 3.2.-.-).
GN	Name=HPSE; Synonyms=HPA;
OS	Gallus gallus (Chicken).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC	Gallus.
ON	NCBI_TaxID=9031;
RN	[1]

RP	NUCLEOTIDE SEQUENCE [MRNA].
RX	MEDLINE=2136959; PubMed=11387326; DOI=10.1074/jbc.M102462200;
RA	Goldschmidt O., Zacharia E., Aingorn H., Guatta-Rangini Z., Atzmon R.,
RA	Michal I., Becker I., Mitrani E., Vlodavsky I.;
RT	"Expression pattern and secretion of human and chicken heparanase are
RT	determined by their signal peptide sequence.";
RL	J. Biol. Chem. 276:29178-29187(2001).
CC	-!- FUNCTION: Endoglycosidase which is a cell surface and
CC	extracellular matrix-degrading enzyme. Cleaves heparan sulfate
CC	proteoglycans (HSPGs) into heparan sulfate side chains and core
CC	proteoglycans (By similarity).
CC	-!- SUBCELLULAR LOCATION: Lysosomal; membrane-associated and secreted
CC	(By similarity).
CC	-!- PTM: N-glycosylated (By similarity).
CC	-!- SIMILARITY: Belongs to the glycosyl hydrolase 79 family.
CC	-----
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC	Distributed under the Creative Commons Attribution-NoDerivs License
CC	-----
DR	EMBL; AY037007; AA82648.1; -, mRNA.
DR	Ensembl; ENSGALG0000011203; Gallus gallus.
DR	InterPro; IPR005199; Glyco_hydro_79_N.
DR	Pfam; PF03662; Glyco_hydro_79n; 1.
KW	Glycoprotein; Hydrolase; Lysosome; Membrane; Signal.
FT	SIGNAL 1 18 Potential.
FT	CHAIN 19 523 Heparanase.
FT	REGION 137 141 Heparin/H5-binding (By similarity).
FT	REGION 250 260 Heparin/H5-binding (By similarity).
FT	ACT SITE 204 204 Proton donor (Potential).
FT	ACT SITE 323 323 Nucleophile (Potential).
FT	CARBOHYD 141 141 N-linked (GlcNAc . .) (Potential).
FT	CARBOHYD 196 196 N-linked (GlcNAc . .) (Potential).
FT	CARBOHYD 436 436 N-linked (GlcNAc . .) (Potential).
FT	CARBOHYD 439 439 N-linked (GlcNAc . .) (Potential).
SQ	SEQUENCE 523 AA; 58386 MW; 8BE0B7B18C9BF81 CRC64;
	/FTID=PRO_0000042259.
Query Match	66.6%; Score 1345.5; DB 1; Length 523;
Best Local Similarity	65.1%; Pred. No. 9,7e-100;
Matches 252, Conservative	63; Mismatches 71; Indels 1; Gaps 1
Qy	1 KKFKNSTYSRSSVDLVTFANCSGLDLIFGNALLRTADLQWNSSNAQLLLDYCSKGYN 60
Db	137 KCHKNTTITRSLDLHTFASSGFRLVFGNLALLRRAGLQWDSSNAQLLYCYCAQRSYN 196
Qy	61 ISWELGNPNSEFLKKADIFNGSOLGEDFIOLHKLL-RKSTFKNAKLYCPDVGPERRKTA 119
Db	197 ISWELGNPNSEFRKKSIGCIDGFGQLGRDFVHLRQLLSQHPLYRHAELYGLDVGQPKKHQT 256
Qy	120 KMUKSFLKAGEVIDSWTHYYLYNGRTATREDFLNPDLVDLFISSVKVFQVVESTREP 179
Db	257 HLLRSFMKGGAIDSVTWHYYLYNGRSATREDFLSPVELDSFATAIHDLVLGIVEATVP 316
Qy	180 KKWLGTSTSAYGCGAPLLSDTFAGGFMWLDKLGLSRMGIEVMNRQVFPGAGNTHLVD 239
Db	317 KKWLGTSTSAYGCGAPQLSNTRYAGFWMLDKGLAARRGIDVVNRQVSFGAGSYHLVDA 376
Qy	240 NFDPPLPYWLSLLPKLVGTVLMASVQSRRKLRVYLHCTNDTPRYKEGDLTLYAIN 299
Db	377 GFKEPLPYWLSLLYKRVLGTVLQAQSVQADARRPVRVYLHCTNPHPKYREGDVTLFALN 436
Qy	300 LHNVTKYLRLPYPFSNKOVDKYLLRPLGPHGLLSKSVOLNGTLTKMVDVDTLPPLMEKPL 359
Db	437 LSNVTQSLQPLKQLWSKVSDQVLLPHGKDLSILREVQLNGELLQWDDETLPALHEMAL 496
Qy	360 RPSSGLPLAFSYSFFVIRNAKVAACI 386
Db	497 APGSTLGLPAFSYGFYVIRNAKAIACI 523
RESULT 11	
Q4SYF6.TETNG	
ID Q4SYF6.TETNG	PRELIMINARY; PRT; 533 AA.

RESULT 11  
Q4SYF6\_TE  
ID Q4SY

AC O4SYF6;  
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
DT 19-JUL-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Chromosome undetermined SCAF12073, whole genome shotgun sequence.  
DE (Fragment).  
GN ORFNames=GSTENG00010356001;  
OS Tetraodon nigroviridis (Green puffer).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
OC Tetraodontoidea; Tetraodontidae; Tetraodon.  
OX NCBI\_TaxID=99883;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15496914; DOI=10.1038/nature03025;  
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,  
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,  
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,  
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,  
RA Biemont C., Skalli Z., Cattolico L., Poulin J., De Berardinis V.,  
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,  
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Boeak S.,  
RA Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,  
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.,  
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals  
RT the early vertebrate proto-karyotype";  
RL Nature 431:946-957(2004).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RG Genoscope; Whitehead Institute Centre for Genome Research;  
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; CAAB01012073; CAF94326.1; -; Genomic\_DNA.  
FT NON TER 1  
FT NON TER 533  
FT NON TER 533  
SQ SEQUENCE 533 AA; 60100 MW; 9B00A7C8780100FF CRC64;  
Query Match 52.5%; Score 1059.5; DB 2; Length 533;  
Best Local Similarity 50.2%; Pred. No. 1.3e-76;  
Matches 213; Conservative 67; Mismatches 103; Indels 41; Gaps 7;  
QY 1 KKFKNSTYSRSDVLYTFANCGLDIFGLNALLRTADLQWNSNAQLLDYCCKGYN 60  
DB 113 KRYRRVKFTETTTVDQLHAFANCSGLDLVFGNALLRTADNRNNSNARSLLRYCEARYH 172  
QY 61 ISWELGNEPNSFLKXADIFINGSOLGDFTLQHLKLRKSTF-KNAKLYGPDVGOPRKRKA 119  
DB 173 MSWELGNEPNSYKKAQRLDGLGOLGDFFTVLKILRESFYRDAGLFGPDVGOPRHHRI 232  
QY 120 KMLKSLFKAGGEVIDSWTHHYLYNGRTATREDFNPDVLDIFISSYQKVFQVVESTRPG 179  
DB 233 DILSGFLQSGAEAVDACTWHYHYLDGREASLEDFDPDVLTLREKIGEVLEHVQVSPG 292  
QY 180 KKWLGETSAYGGGAPLLSDTFAAGPMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDE 239  
DB 293 KPWLGETSAYGGGAAGLSDTFFVAGPMWLDKGLAATGLGLVWRQVLGAGSYHLMD 352  
QY 240 NFDPLP-----DYLWLSLFLKVLGTFVLMA-----SVQSKRKLRYLHCTNTDN--- 285  
DB 353 NLDPLPRSGLLLDQYWLSSLYKXLVGQEVLEKTRHTPTGPAGSER--VRLYLHCANKQRCSL 410  
QY 286 -----PRYKEGDLTLVAINLHNVTYLRPLPYFPFSKNQVDKYL-- 323

Db 411 LQFLSVRKQRKEARFSLVSLCSYRSRGAATLMSMNLKQPARISLPRILSSSTVEAFVLES 470  
QY 324 -RPLGPHGLLSKSVOLNGLTLKMYDDQTLPLMEKPLRPGSSGLGPAFYSYSFVIRNAKV 382  
Db 471 EQP-GEGLSRVAVKLNGRVLRMVDDTFPELEGSRLPAAEHLQLPAYSLAFVFTDAQA 529  
QY 383 AACI 386  
Db 530 AGCV 533  
RESULT 12  
HPSE2\_HUMAN STANDARD; PRT: 592 AA.  
ID Q8WHQ2; Q5VUH4; Q5VUH5; Q5VUH6; Q6WQ1; Q9HB37; Q9HB38; Q9HB39;  
DT 25-OCT-2005, integrated into UniProtKB/Swiss-Prot.  
DT 25-OCT-2005, sequence version 2.  
DT 07-MAR-2006, entry version 16.  
DE Heparanase-2 (EC 3.2.-.-) (Hpa2).  
GN Name=HPSE2; Synonyms=HPA2;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA] (ISOFORMS 1; 3 AND 4), TISSUE SPECIFICITY,  
RP AND SUBCELLULAR LOCATION.  
RC TISSUE=Heart;  
RX MEDLINE=20483645; PubMed=11027606; DOI=10.1006/bbrc.2000.3586;  
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,  
RA Hircok M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;  
RT "Cloning and expression profiling of Hpa2, a novel mammalian  
RL heparanase family member";  
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [MRNA] (ISOFORMS 1 AND 2).  
RC TISSUE=Prostate;  
RA Legoux P., Legoux R., O'Brien D., Salome M.;  
RN Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA], AND VARIANT TYR-579.  
RX PubMed=15164054; DOI=10.1038/nature02462;  
RA Deloukas P., Earthworm M.E., Grafham D.V., Rubinfeld M., French L.,  
RA Steward C.A., Sims S.K., Jones M.C., Searle S., Scott C., Howe K.,  
RA Hunt S.E., Andrews T.D., Gilbert J.G.R., Swarbreck D., Ashurst J.L.,  
RA Taylor A., Battles J., Bird C.P., Ainscough R., Almeida J.P.,  
RA Ashwell R.I.S., Ambrose K.D., Babbage A.K., Baguley C.L., Bailey J.,  
RA Banerjee R., Bates K., Beasley H., Bray-Allen S., Brown A.J.,  
RA Brown J.Y., Burford D.C., Burrill W., Burton J., Cahill P., Camire D.,  
RA Carter N.P., Chapman J.C., Clark S.Y., Clarke G., Clee C.M., Clegg S.,  
RA Corby N., Coulson A., Dhami P., Dutta I., Dunn M., Faulkner L.,  
RA Frankish A., Frankland J.A., Garner P., Garnett J., Gribble S.,  
RA Griffiths C., Grocock R., Gustafson E., Hammond S., Harley J.L.,  
RA Hart E., Heath P.D., Ho T.P., Hopkins B., Horne J., Howden P.J.,  
RA Huckle E., Hynds C., Johnson C., Johnson D., Kana A., Kay M.,  
RA Kimberley A.M., Kershaw J.K., Kokkinaki M., Laird G.K., Lawlor S.,  
RA Lee H.M., Leongamornlert D.A., Laird G., Lloyd C., Lloyd D.M.,  
RA Loveland J., Lovell J., McLaren S., McLay K.E., McMurray A.,  
RA Madhreggi-Mohammadi M., Matthews L., Milne S., Nickerson T.,  
RA Nguyen M., Overton-Larty E., Palmer S.A., Pearce A.V., Peck A.I.,  
RA Pelan S., Phillimore B., Porter K., Rice C.M., Rogosin A., Ross M.T.,  
RA Sarafidou T., Sehra H.K., Showkeen R., Skuce C.D., Smith M.,  
RA Standing L., Sycamore N., Tester J., Thorpe A., Torcaso W.,  
RA Tracey A., Tromans A., Tsolas J., Wall M., Walsh J., Wang H.,  
RA Weinstock K., West A.P., Willey D.L., Whitehead S.L., Wilming L.,  
RA Wray P.W., Young L., Chen Y., Lovering R.C., Moschonas N.K.,  
RA Siebert R., Fectel K., Bentley D., Durbin R., Hubbard T.,  
RA Doucette-Stamm L., Beck S., Smith D.R., Rogers J.;  
RT "The DNA sequence and comparative analysis of human chromosome 10";  
RL Nature 429:375-381(2004).  
CC -!- FUNCTION: Endoglycosidase which is a cell surface and  
CC extracellular matrix-degrading enzyme. Cleaves heparan sulfate





```

Db 370 TYPGKIMLEGVVTTAGGTTNNLSDSYAAGFLWNTLGMLANOGIDVIRHSFFDHGYN 429
QY 235 HLVDENFDPLPDYWLKLLFKLVGTVKVLMSVQSGSKRR-----KLRVYLHCTNTDN 285
Db 430 HLVDQNFNPLPDYWLKLLFKLVGTVKVLMSVQSGSKRR-----KLRVYLHCTNTDN 489
QY 286 PRYKEDGLTIYALNHNVTYKYLRLPYFPFSNKQVDKYLRLPLGPHGLSKSVQNLGLTKM 345
Db 490 HNYVRSITLFIINLHRSRKKIKLAGTLRKLVHQYLLQPYQSGGLSKSVQNLGLTKM 549
QY 346 VDDQTLPLMEKPLRPGSSGLGPAFSYFFVIRNAKVAAC 385
Db 550 VDDGTLPELKPRLRAGRTLIVPVTMGFFVVKVNNALAC 589

RESULT 13
Q2M1H9_HUMAN
ID Q2M1H9_HUMAN PRELIMINARY; PRT; 592 AA.
AC Q2M1H9
DT 21-FEB-2006, integrated into UniProtKB/TrEMBL.
DT 21-FEB-2006, sequence version 1.
DE Hepatanase 2.
GN Name=HPSE2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
NUCLEOTIDE SEQUENCE.
RP TISSUE=PCR rescued clones;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Shenmen C.M., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Schaefer C.F., Bhat N.K.,
RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
RA Diatchenko L., Marushina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Tohiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
NUCLEOTIDE SEQUENCE.
RP TISSUE=PCR rescued clones;
RC NIH MGC Project;
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; BC112356; AAI12357.1; -; mRNA.
CC SEQUENCE 592 AA; 66610 MW; 94689E1C2A74359F CRC64;

Query Match 48.4%; Score 978; DB 2; Length 592;
Best Local Similarity 48.8%; Pred. No. 5,7e-70;
Matches 195; Conservative 60; Mismatches 129; Indels 16; Gaps 4;

QY 1 KKFKNSTYSR-----SSVDVLYTFANCSGDLDFGLNALLRTADLQWNSNAQLLDYCS 55
Db 191 EQFSN-TYSNLIILARSGLKYNFADCSGLHLFALNALRNPNNWNSSALSLLKYSA 249

```

```

QY 56 SKGYNISWELGNEPNSFLKKADIFINGSQLEDGFQLHKLRLK-STFKNAKLYGPDVGOP 114
Db 250 SKKYNISWELGNEPNNYRTMGRAVNSQGLKDYIQLKSLQPIRIYSRASLYGPNIGRP 309
QY 115 RRKTAKMLKSLKAGGEVIDSVTHHHYHNGRTATREDPLNDPDLIDIFISSVQKVFQVVE 174
Db 310 RKNVIALLDGFMKVGASTVDATWQHCYIDGRVVKVMDFLKTRLDLTDLSQIRKIOKVN 369
QY 175 STRFGKVKWLGETSSAYGGAPLLSDTFAAGPMWLDKGLSARMGIEVVMRQVFFGAGNY 234
Db 370 TYPGKIMLEGVVTTAGGTTNNLSDSYAAGFLWNTLGMLANOGIDVIRHSFFDHGYN 429
QY 235 HLVDENFDPLPDYWLKLLFKLVGTVKVLMSVQSGSKRR-----KLRVYLHCTNTDN 285
Db 430 HLVDQNFNPLPDYWLKLLFKLVGTVKVLMSVQSGSKRR-----KLRVYLHCTNTDN 489
QY 286 PRYKEDGLTIYALNHNVTYKYLRLPYFPFSNKQVDKYLRLPLGPHGLSKSVQNLGLTKM 345
Db 490 HNYVRSITLFIINLHRSRKKIKLAGTLRKLVHQYLLQPYQSGGLSKSVQNLGLTKM 549
QY 346 VDDQTLPLMEKPLRPGSSGLGPAFSYFFVIRNAKVAAC 385
Db 550 VDDGTLPELKPRLRAGRTLIVPVTMGFFVVKVNNALAC 589

RESULT 14
Q4TB80_TETNG
ID Q4TB80_TETNG PRELIMINARY; PRT; 597 AA.
AC Q4TB80
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DE Chromosome 17 SCAP7180, whole genome shotgun sequence. (Fragment).
GN ORFNames=GSTENG0003868001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
NUCLEOTIDE SEQUENCE.
RP PubMed=15496914; DOI=10.1038/nature03025;
RX Jallou O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Basilya C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthonard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poullain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J., Robinson-Rechavi M.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Scallius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
[2]
NUCLEOTIDE SEQUENCE.
RP Genoscope; Whitehead Institute Centre for Genome Research;
RG Genoscope (FEB-2004) to the EMBL/GenBank/DBJ databases.
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; CAAB01007180; CAF89852.1; -; Genomic_DNA.
CC NON TER 597 597
CC SEQUENCE 597 AA; 67127 MW; 83B46AC5B727A8FE CRC64;

```



Query Match 43.3%; Score 874; DB 2; Length 597;  
Best Local Similarity 45.2%; Pred. No. 1.4e-61;  
Matches 180; Conservative 61; Mismatches 141; Indels 16; Gaps 5;  
CC 3 FKNSTYSRSVDVLYTFANCSGLDLIFGLNALLRTADLQWSSNAQLLDYCSSKGYNIS 62  
DB 200 FSHSRIAR-SLDKLYNFADACAGLHILGLNALHRNPDHSWNTSTLLKYSAGKYNIS 258  
QY 63 WELGNEPNSFLKXKADIFINGSQGLDFIOLHKLRLK-STFKNAKLYGPDVGPQRRKTAKM 121  
DB 259 WELGNEPNAYSRVMGHAVNSQAQDYTKURLQSVRYSRQLYGPNAGRPRKNALL 318  
QY 122 LKSFKAGGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVQFQVVESTREGKK 181  
DB 319 LDEFMKTVGVDAVTHQHYMDGRKKVEDEFLKTRLLDITLQLSKVTQVNVHTTSGKK 378  
QY 182 VWLGETSSAYGGAPLLSDTFAAGFWMLDKLGUSARMGIEVVMR-----QVFFAGNTHLV 237  
DB 379 VWLGGGLGPAWTGMSNLSDTFAAGFLVNTLGMAMQGDIVLRRQAVQSBHTNKQSVLF 438  
QY 238 DENEDP-LPDYWLFLFKKLVGTVKVLMAVQGSKR-----KLRVVLHCTNTDNP 287  
DB 439 LQMFVPSFPDYWFSLVFKLVGPKVLAVRQVAGLQRPQGRVTRDKLRIYARCTSYSNHN 498  
QY 288 YKEGDLTLYAINLHNVTKYLRPYPESNKQVDKYLLRPLGPHGLLSKSVOLNGLTLQWVD 347  
DB 499 YVRGSIITIIILNRSRKKIKLAGLIRNNIVHQLQPYGADGLRAKHVQNLSEKLLMAD 558  
QY 348 DQTLPLMEKPLRPGSSGLPAPFSYFFVIRNAKVAAC 385  
DB 559 NETFPELKPRTLRAGRTIAMPPTGIFYVYIKINAYAC 596

## RESULT 15

Q4TCG8 TETNG PRELIMINARY; PRT; 255 AA.  
AC Q4TCG8;  
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
DT 19-JUL-2005, sequence version 1.  
DE Chromosome undetermined SCAF3783, whole genome shotgun sequence.  
DE (Fragment).  
GN ORFNames=GSTENG0001168001;  
OS Tetraodon nigroviridis (Green puffer).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
OC Tetraodontidae; Tetraodontidae; Tetraodon.  
OX NCBI\_TaxID=99883;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15496914; DOI=10.1038/nature03025;  
RA Jallion O., Aury J.-M., Brunet P., Petit J.-L., Stange-Thomann N.,  
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,  
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,  
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
RA Anthouard V., Jubin C., Castelli V., Poulain M., Vacherie B.,  
RA Bienmont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,  
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,  
RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,  
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,  
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;  
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals  
the early vertebrate proto-karyotype.";  
RL Nature 431:946-957(2004).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RG Genoscope; Whitehead Institute Centre for Genome Research;  
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an

CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; CAAB01003783; CAF88054.1; -; Genomic\_DNA.  
FT NON TER 1 1  
FT NON TER 255 255  
SQ SEQUENCE 255 AA; 28562 MW; 07F542A9C755E3F0 CRC64;  
Query Match 36.8%; Score 742.5; DB 2; Length 255;  
Best Local Similarity 56.7%; Pred. No. 1.9e-51;  
Matches 144; Conservative 37; Mismatches 56; Indels 17; Gaps 3;  
QY 9 SRSSVDVLYTFANCSGLDLIFGLNALLRTADLQWSSNAQLLDYCSSKGYNISWELGNE 68  
DB 1 SETTVDDQLHAFANCSGLDLVFGLNALLRTADNRWSSNARSLLRYCEARRYHMSWELGNE 60  
QY 69 PNSFLKKADIFINGSQGLDFIOLHKLRLKSTF-KNAKLYGPDVGPQRRKTAKMLKSLK 127  
DB 61 PNSYEKKAGLRDGRQLGEDFTVLRKILRESRFYRDAGLFGPDVGPQPRDHRIDILSGFLQ 120  
QY 128 AGGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVQFQVVESTREGKKVWLGET 187  
DB 121 SGAEAVDACTWHHYLDGREASLEDFLDPDVLDITREKIGEVLEEVHQSVPKPVWLGET 180  
QY 188 SSAYGGG-----APLLSDTFAAG-FMWLDKLGLSARMGIEVVMQVFFGA 231  
DB 181 SSATGAEPGRCTHSSQDSCEFAFRSDQAPLGTFRWLDKLGLAATLGLVLMQVFLGA 240  
QY 232 GNYHLVDENFDPLP 245  
DB 241 GSTHLMDDNLDPLP 254

Search completed: June 5, 2006, 12:20:13

Job time : 96.7733 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:01:06 ; Search time 76.9552 Seconds  
(without alignments)  
2293.354 Million cell updates/sec

Title: US-10-645-659A-1  
Perfect score: 2020  
Sequence: 1 KFKKSTYSRSSVDVLYTFA.....LPAFSYSFFVIRNAKVAACI 386

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_8:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*  
9: Geneseqp2005s:\*  
10: Geneseqp2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2020	100.0	386	8	Adr88207 Human mat
2	2020	100.0	386	8	Adt78174 45kDa sub
3	2020	100.0	386	9	Ady27057 Heparanas
4	2020	100.0	386	9	Adz18995 Human hep
5	2020	100.0	386	9	Aea42423 Human mat
6	2020	100.0	460	9	Agv27061 Heparanas
7	2020	100.0	422	9	Adz18996 Hep106 co
8	2020	100.0	495	9	Adz18999 Hep109 co
9	2020	100.0	501	9	Adz19000 HepG3 co
10	2020	100.0	507	9	Adz19005 HepG6 co
11	2020	100.0	508	9	Ady27058 Human ina
12	2020	100.0	526	9	Adz19006 HepHyalur
13	2020	100.0	527	5	Abb07815 Chicken s
14	2020	100.0	527	7	Adw02018 Chimeric
15	2020	100.0	527	9	Adz19004 HepG4 co
16	2020	100.0	543	2	Aay17082 Human hep
17	2020	100.0	543	4	Aab86206 Human hep
18	2020	100.0	543	7	Adb18950 Human dis
19	2020	100.0	543	8	Adk52086 Human ato
20	2020	100.0	543	8	Adm48759 Human hpa
21	2020	100.0	543	8	Adn05074 Antipsori
22	2020	100.0	543	8	Adn04902 Antipsori
23	2020	100.0	543	8	Adq80372 Heparanas

24	2020	100.0	543	8	ADR88210	Human pre
25	2020	100.0	543	8	ADP25079	PRO poly
26	2020	100.0	543	8	ADT78177	Human hep
27	2020	100.0	543	9	ADY27036	Human hep
28	2020	100.0	543	9	AEA42426	Human hep
29	2020	100.0	545	6	ABP56822	Human hep
30	2020	100.0	545	7	ADE16012	G-coupled
31	2020	100.0	545	8	ADL93951	Human G-c
32	2020	100.0	556	9	ADZ19010	Heparanas
33	2020	100.0	570	9	ADZ19008	Heparanas
34	2020	100.0	588	2	AAZ30124	A human p
35	2017	99.9	530	2	AAZ34173	Human pre
36	2017	99.9	543	2	AAV02345	A human h
37	2017	99.9	543	3	AAV57590	Human hep
38	2017	99.9	543	3	AAB08849	Amino aci
39	2017	99.9	543	3	AAV52990	Human hep
40	2017	99.9	543	4	AAV97635	Human hep
41	2017	99.9	543	5	ABB07813	Human hep
42	2017	99.9	543	7	ADG88800	Human hpa
43	2017	99.9	543	8	ADL16379	Human hep
44	2017	99.9	543	8	ADM48716	Human hpa
45	2017	99.9	543	9	AEA42466	Human hep

ALIGNMENTS

RESULT 1  
ADR88207  
ID ADR88207 standard; protein; 386 AA.  
XX  
AC ADR88207;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Human mature heparanase 45 kDa major subunit.

Targeted drug delivery ; inflammatory disorder; wound; scar;  
vasculopathy; autoimmune disorder; cancer; angiogenesis;  
metastatic disease; atherosclerosis; restenosis; aneurysm; solid cancer;  
non-solid cancer; haematopoietic malignancy ; lymphocytic leukaemia;  
myelogenous leukaemia; Hodgkin's disease; multiple myeloma;  
haemangiosarcoma; Kaposi's sarcoma; human ; heparanase; enzyme.

Homo sapiens.

US2004170631-A1.

02-SEP-2004.

28-NOV-2003; 2003US-00722502.

02-SEP-1997; 97US-00922170.

01-MAY-1998; 98US-00071739.

14-NOV-1998; 98US-00186200.

19-FEB-2003; 2003US-00368044.

22-AUG-2003; 2003US-00645659.

(YACO/) YACOBY-ZEEVI O.

(PERE/) PERETZ T.

(MIRO/) MIRON D.

(SHLO/) SHLOMI Y.

(PECK/) PECKER I.

(AYAL/) AYAL-HERSHKOVITZ M.

(FEIN/) FEINSTEIN E.

(VAGEL/) VAN GELDER J M.

(VLOD/) VLODAVSKY I.

(FRIE/) FRIEDMANN Y.

Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
Ayal-Hershkovitz M, Feinstein E, Van Gelder JM, Vlodavsky I;  
Friedmann Y;

DR WPI; 2004-625084/60.  
XX Targeted drug delivery to a heparanase-expressing tissue of a patient,  
PT useful for treating heparanase-associated conditions such as inflammation  
PT or cancer, comprises administering a drug and an anti-heparanase antibody  
PT complex.  
XX  
PS Claim 2; SEQ ID NO 1; 58pp; English.  
XX  
CC The invention relates to a method of targeted drug delivery to a tissue  
CC of a patient, the tissue expressing heparanase. The method comprises  
CC providing a complex of a drug directly or indirectly linked to an anti-  
CC heparanase antibody, and administering the complex to the patient. In the  
CC targeted drug delivery, the antibody comprises an epitope of a heparanase  
CC capable of specifically binding to at least one epitope of a heparanase  
CC protein. The composition and methods of the invention are useful for  
CC diagnosing, preventing or treating conditions associated with heparanase  
CC catalytic activity (e.g. an inflammatory disorder, wound, scar,  
CC vasculopathy, an autoimmune disorder, cancer, angiogenesis, cell  
CC proliferation, invasion of circulating tumour cells and metastatic  
CC disease), for purifying heparanase, or for developing drugs for those  
CC heparanase-associated conditions. The vasculopathy is atherosclerosis,  
CC restenosis or aneurysm. The cancerous condition is a solid cancer or a  
CC non-solid cancer. The non-solid cancer is a haematopoietic malignancy  
CC selected from acute lymphocytic leukaemia (ALL), acute myelogenous  
CC leukaemia, chronic lymphocytic leukaemia (CLL), chronic myelogenous  
CC leukaemia (CML), myelodysplastic syndrome (MDS), mast cell leukaemia,  
CC Hodgkin's disease, non-Hodgkin's lymphomas, Burkitt's lymphoma and  
CC multiple myeloma. The solid cancer is selected from tumours in lip and  
CC oral cavity, pharynx, larynx, paranasal sinuses, major salivary glands,  
CC thyroid gland, oesophagus, stomach, small intestine, colon, colorectum,  
CC anal canal, liver, gallbladder, extrahepatic bile ducts, ampulla of  
CC Vater, exocrine pancreas, lung, pleural mesothelioma, soft tissue  
CC sarcoma, carcinoma and malignant melanoma of the skin, breast, vulva,  
CC vagina, cervix uteri, ovary, fallopian tube, gestational trophoblastic  
CC tumours, penis, prostate, testis, kidney, renal pelvis, ureter, urinary  
CC bladder, urethra, carcinoma of the eyelid, carcinoma of the conjunctiva,  
CC malignant melanoma of the conjunctiva, malignant melanoma of the uvea,  
CC retinoblastoma, carcinoma of the lacrimal gland, sarcoma of the orbit,  
CC brain, spinal cord, vascular system, haemangiosarcoma and Kaposi's  
CC sarcoma. The present sequence is the 45 kDa major subunit of human mature  
CC heparanase.  
XX  
SQ Sequence 386 AA;  
Query Match 100.0%; Score 2020; DB 8; Length 386;  
Best Local Similarity 100.0%; Pred. No. 4.9e-202;  
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKFKNSTYGRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60  
DB 1 KKFKNSTYGRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60  
QY 61 ISWELGNPNFLKKADIFINGSQLEDFTQLHKLKSTFKNAKLYGPDVGQPRRTAK 120  
DB 61 ISWELGNPNFLKKADIFINGSQLEDFTQLHKLKSTFKNAKLYGPDVGQPRRTAK 120  
QY 121 MLKSLFKAGGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSQVKVQVVESTPGK 180  
DB 121 MLKSLFKAGGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSQVKVQVVESTPGK 180  
QY 181 KYWLGETTSAYGGAPLLSDTFAAGFMWLDKGLSARMGLEVMVRQVFFCAGNYHLVDEN 240  
DB 181 KYWLGETTSAYGGAPLLSDTFAAGFMWLDKGLSARMGLEVMVRQVFFCAGNYHLVDEN 240  
QY 241 FDPFLDYWLSLLFKLVGTVKVLMSYVQSGKRRKRLRVYLHCTNTDNPYKRGDITLYAINL 300  
DB 241 FDPFLDYWLSLLFKLVGTVKVLMSYVQSGKRRKRLRVYLHCTNTDNPYKRGDITLYAINL 300  
QY 301 HNVTKYRLPYFPFNQVDKYLRLPGHGLLSKSVQLNGLTLKQVDDQTLPLMEKPLR 360  
DB 301 HNVTKYRLPYFPFNQVDKYLRLPGHGLLSKSVQLNGLTLKQVDDQTLPLMEKPLR 360

QY 361 PGSSGLPAFSYSFFVIRNAKVAACI 386  
DB 361 PGSSGLPAFSYSFFVIRNAKVAACI 386  
RESULT 2  
ADT78174  
ID ADT78174 standard; protein; 386 AA.  
XX AC  
XX ADT78174;  
XX AC  
XX 13-JAN-2005 (first entry)  
XX DE 45kDa subunit of mature processed human heparanase dimer.  
XX KW Antibody; epitope; heparanase; pathological condition; angiogenesis;  
KW cell proliferation; cancerous condition; tumour cell invasion;  
KW metastatic disease; heparanase-related disorder; inflammatory disorder;  
KW wound; scar; vasculopathy; autoimmune condition; renal disease;  
KW cytostatic; antiinflammatory; vulnery; antiarteriosclerotic;  
KW vasotropic; immunosuppressive; nephrotropic; antidiabetic; human.  
XX OS Homo sapiens.  
XX US2004213789-A1.  
XX 28-OCT-2004.  
XX 22-AUG-2003; 2003US-00645659.  
XX 02-SEP-1997; 97US-00922170.  
PR 01-MAY-1998; 98US-00071739.  
PR 04-NOV-1998; 98US-00186200.  
PR 19-FEB-2003; 2003US-00368044.  
XX (YACO/) YACOBY-ZEEVI O.  
PA (PERE/) PERETZ T.  
PA (MIRO/) MIRON D.  
PA (SHLO/) SHLOMI Y.  
PA (PECK/) PECKER I.  
PA (AYAL/) AYAL-HERSHKOVITZ M.  
PA (FEIN/) FEINSTEIN E.  
PA (GELD/) GELDER J M V.  
PA (VLOD/) VLODAVSKY I.  
PA (FRIE/) FRIEDMANN Y.  
XX Yacoby-Zeevi O, Peretz T, Miron D, Shlomi Y, Pecker I;  
PI Ayal-Hershkovitz M, Feinstein E, Gelder JMW, Vlodavsky I;  
PI Friedmann Y;  
XX WPI; 2004-774790/76.  
XX  
XX New neutralizing monoclonal anti-heparanase antibodies, useful for  
PT detecting, treating or preventing cancer, inflammatory or autoimmune  
PT disorder, wound, atherosclerosis, restenosis, or diabetic nephropathy.  
XX  
PS Claim 5; SEQ ID NO 1; 68pp; English.  
XX  
CC The invention relates to an isolated antibody or antibody portion capable  
CC of specifically binding to or elicited by at least one epitope of a  
CC heparanase protein, where the heparanase protein is at least 60%  
CC homologous to any of the 6 sequences given as SEQ ID NOS: 1-5 or 11, and  
CC where at least one epitope comprises a sequence at least 70% homologous  
CC to any of the sequences given as SEQ ID NOS: 6-10. Also disclosed are (a)  
CC a hybridoma cell line comprising a cell line for producing the monoclonal  
CC antibody, (b) a method for detecting, treating or preventing a  
CC pathological condition or a heparanase-related disorder or condition in a  
CC subject, (c) a method for monitoring the state of a heparanase-related  
CC disorder or condition in a subject, and (d) a pharmaceutical composition  
CC comprising the isolated anti-heparanase antibody or antibody portion and  
CC a pharmaceutical carrier. The antibody, methods, and composition are  
CC useful for detecting, treating, preventing or monitoring a pathological  
CC condition, e.g. angiogenesis, cell proliferation, a cancerous condition

CC (blood, breast, bladder, rectum, stomach, cervix, ovarian, colon, renal,  
 CC or prostate cancer), minor cell proliferation, invasion of circulating  
 CC tumour cells, or a metastatic disease, or a heparanase-related disorder  
 CC or condition, e.g. an inflammatory disorder, wound, scar, vasculopathy  
 CC (atherosclerosis, restenosis, or aneurysm), autoimmune condition, or  
 CC renal disease or disorder (diabetic nephropathy, glomerulosclerosis,  
 CC nephrotic syndrome, minimal change nephrotic syndrome, or a renal cell  
 CC carcinoma) in a mammal. This sequence represents the 45kDa subunit of  
 CC mature processed human heparanase dimer.

XX  
 SQ Sequence 386 AA;  
 Query Match 100.0%; Score 2020; DB 8; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-202;  
 Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60  
 DB 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60  
 QY 61 ISWELGNEPNSFLKADIFINGSQGLDFIQLHKLKSTFKNAKLYGPDVGQPRRTAK 120  
 DB 61 ISWELGNEPNSFLKADIFINGSQGLDFIQLHKLKSTFKNAKLYGPDVGQPRRTAK 120  
 QY 121 MLKSFLLKAGGEVDSVTWHYLLNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGK 180  
 DB 121 MLKSFLLKAGGEVDSVTWHYLLNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGK 180  
 QY 181 KWLGTSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240  
 DB 181 KWLGTSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240  
 QY 241 FDPPLDYWLSLLFKKLVGTVKVMASVQSGSKRRKRLRVYLHCTNTDNPYKEGDLTYAINL 300  
 DB 241 FDPPLDYWLSLLFKKLVGTVKVMASVQSGSKRRKRLRVYLHCTNTDNPYKEGDLTYAINL 300  
 QY 301 HNTVKYLRPLPYPSNKKQVDKYLRLPLGPHGLLSKSVQNLGLTLKQVDDQTLPLMEKPLR 360  
 DB 301 HNTVKYLRPLPYPSNKKQVDKYLRLPLGPHGLLSKSVQNLGLTLKQVDDQTLPLMEKPLR 360  
 QY 361 PGSSGLGPAFSYFFVIRNAKVAACI 386  
 DB 361 PGSSGLGPAFSYFFVIRNAKVAACI 386

RESULT 3

ADY27057

ID ADY27057 standard; protein; 386 AA.

AC ADY27057;

XX

XX

DT 05-MAY-2005 (first entry)

XX

XX

DE Heparanase inhibitor protein #1.

XX

KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;

KW neurological disease; viral infection; infection; cytostatic;

KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;

XX heparanase modulator; enzyme purification.

XX

OS Homo sapiens.

XX

PN WO2005016227-A2.

XX

PD 24-FEB-2005.

XX

PF 12-AUG-2004; 2004WO-IL000744.

XX

PR 14-AUG-2003; 2003US-0494800P.

XX

PR 12-JAN-2004; 2004US-0535492P.

XX

PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.

XX

PI Van-Gelder JM, Miron D;  
 DR WPI; 2005-182203/19.  
 XX  
 PT Regulating heparanase activity, useful for treating heparanase-associated  
 PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
 PT neurological diseases or viral diseases) comprises modulating heparanase  
 PT activation.  
 XX  
 PS Claim 55; SEQ ID NO 33; 211pp; English.

XX  
 CC The invention relates to a method of regulating heparanase activity in a  
 CC tissue or regulating a biological process depending at least in part on  
 CC heparanase activity comprising modulating heparanase activation. The  
 CC invention also relates to methods of treating a heparanase- or heparin  
 CC binding protein-associated disease or disorder in a subject, a  
 CC pharmaceutical composition for use in the treatment of a heparanase-  
 CC associated disease or disorder comprising a therapeutic amount of an  
 CC agent capable of modulating heparanase activation and a pharmaceutical  
 CC carrier or diluent, a method of identifying a protease activator of  
 CC heparanase, a protease substrate mimetic comprising a peptide  
 CC representing a subset or all substrate residues or cleavage sites of  
 CC human heparanase or an equivalent non-human heparanase, a method of  
 CC producing active heparanase and a method of modulating an adhesion  
 CC activity of heparanase. The composition and methods are useful for  
 CC modulating heparanase activation and for treating heparanase-associated  
 CC diseases or disorders such as cancer, inflammation, cardiovascular  
 CC diseases, neurological diseases or viral infections. This sequence  
 CC represents a heparanase inhibitor protein used in the scope of the  
 CC invention.

XX SQ Sequence 386 AA;

Query Match 100.0%; Score 2020; DB 9; Length 386;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-202;  
 Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60  
 DB 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60  
 QY 61 ISWELGNEPNSFLKADIFINGSQGLDFIQLHKLKSTFKNAKLYGPDVGQPRRTAK 120  
 DB 61 ISWELGNEPNSFLKADIFINGSQGLDFIQLHKLKSTFKNAKLYGPDVGQPRRTAK 120  
 QY 121 MLKSFLLKAGGEVDSVTWHYLLNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGK 180  
 DB 121 MLKSFLLKAGGEVDSVTWHYLLNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGK 180  
 QY 181 KWLGTSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240  
 DB 181 KWLGTSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240  
 QY 241 FDPPLDYWLSLLFKKLVGTVKVMASVQSGSKRRKRLRVYLHCTNTDNPYKEGDLTYAINL 300  
 DB 241 FDPPLDYWLSLLFKKLVGTVKVMASVQSGSKRRKRLRVYLHCTNTDNPYKEGDLTYAINL 300  
 QY 301 HNTVKYLRPLPYPSNKKQVDKYLRLPLGPHGLLSKSVQNLGLTLKQVDDQTLPLMEKPLR 360  
 DB 301 HNTVKYLRPLPYPSNKKQVDKYLRLPLGPHGLLSKSVQNLGLTLKQVDDQTLPLMEKPLR 360  
 QY 361 PGSSGLGPAFSYFFVIRNAKVAACI 386  
 DB 361 PGSSGLGPAFSYFFVIRNAKVAACI 386

RESULT 4

ADZ18995

ID ADZ18995 standard; protein; 386 AA.

XX

AC ADZ18995;

XX

DT 16-JUN-2005 (first entry)

XX Human heparanase consensus cleavage site #2.  
DE  
XX Enzyme engineering; heparanase; metastasis; autoimmune disease;  
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
KW immunosuppressive; enzyme.  
XX Homo sapiens.  
OS  
XX WO2005030962-A1.  
PN  
XX 07-APR-2005.  
PD  
XX 17-SEP-2004; 2004WO-EP010517.  
PF  
XX 26-SEP-2003; 2003US-0506479P.  
PR  
XX 20-JAN-2004; 2004US-0537729P.  
PR  
XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.  
PA  
XX Lahm A, Nardella C, Pallaro M, Steinkuhler C;  
PI  
XX WPI; 2005-273382/28.  
DR  
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
PT nucleotide sequence that encodes mammalian heparanase protein and has two  
PT consensus cleavage sites located between specific nucleotide encoding  
PT residues.  
XX  
XX Disclosure; SEQ ID NO 16; 65pp; English.  
PS  
XX The invention relates to a synthetic nucleic acid molecule that encodes  
CC mammalian heparanase protein, where the nucleic acid comprises two  
CC consensus cleavage sites recognized by endoproteinase. The sequences are  
CC useful for expressing mammalian heparanase in non-mammalian cells and in  
CC inhibitor screening assays for the development of therapeutics or  
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
CC and/or inflammation. This sequence represents a human heparanase  
CC consensus cleavage site used in the scope of the invention.  
XX  
XX Sequence 386 AA;  
SQ  
Query Match 100.0%; Score 2020; DB 9; Length 386;  
Best Local Similarity 100.0%; Pred. No. 4.9e-202;  
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKFKNSTYSSRSDVLYTFANCSSGLDLIFGLNALLRTADLQWSSNAQLLDYCSSKGYN 60  
Db 1 KKFKNSTYSSRSDVLYTFANCSSGLDLIFGLNALLRTADLQWSSNAQLLDYCSSKGYN 60  
QY 61 ISWELGNEPNSFLKKADIFINGSOLGEDFTQLHKLKRSFTFNKALYGPDPVGGPRRTAK 120  
Db 61 ISWELGNEPNSFLKKADIFINGSOLGEDFTQLHKLKRSFTFNKALYGPDPVGGPRRTAK 120  
QY 121 MLKSFLLKAGEVDSVTWHYLYNGRTATREDFLNPDVLDIFISSQKVFQVVESTPRGK 180  
Db 121 MLKSFLLKAGEVDSVTWHYLYNGRTATREDFLNPDVLDIFISSQKVFQVVESTPRGK 180  
QY 181 KVMGLTSSAYGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240  
Db 181 KVMGLTSSAYGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240  
QY 241 FDPDPDWLSLFLFKLVGTVKVLMAVQGSRRKRLVYLHCTNTDNPYKEGDLTLVAINL 300  
Db 241 FDPDPDWLSLFLFKLVGTVKVLMAVQGSRRKRLVYLHCTNTDNPYKEGDLTLVAINL 300  
QY 301 HNVTKYLRPLPYPSNKKVDKLLRLPLGPHGLSKSVQLNGLTLKVVDDQTLPLMEKPLR 360  
Db 301 HNVTKYLRPLPYPSNKKVDKLLRLPLGPHGLSKSVQLNGLTLKVVDDQTLPLMEKPLR 360  
QY 361 PGSSGLGLPAFSYFFVIRNAKVAACI 386  
Db 361 PGSSGLGLPAFSYFFVIRNAKVAACI 386

RESULT 5  
AEA42423  
ID AEA42423 standard; protein; 386 AA.  
XX  
AC AEA42423;  
XX  
XX 28-JUL-2005 (first entry)  
XX Human mature heparanase dimer 45 kDa subunit SEQ ID NO:1.  
DE  
XX antibody; heparanase; antiinflammatory; vulnery; immunosuppressive;  
KW antiangiogenic; cytostatic; antiarteriosclerotic; vasotropic;  
KW inflammation; wound healing; scarring; vasculopathy; autoimmune disease;  
KW angiogenesis disorder; cancer; tumor; metastasis.  
XX  
OS Homo sapiens.  
XX  
XX AU2004201462-A1.  
PN  
XX 06-MAY-2004.  
PD  
XX 08-APR-2004; 2004AU-00201462.  
PF  
XX 08-APR-2004; 2004AU-00201462.  
PR  
XX (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.  
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV LTD.  
XX  
XX Vlodavsky I, Pecker I, Mimon M, Gilboa A, Miron D, Moskowitz H;  
PI Shlomi Y, Peleg Y, Yacoby-Zeevi O, Ayal-Hershkovitz M, Ben-Artzi H;  
PI Feinstein E;  
XX  
DR WPI; 2005-173343/19.  
XX  
XX Novel isolated antibody capable of specifically binding to epitope of  
PT heparanase protein, useful for preventing and treating heparanase-related  
PT disorder such as inflammatory disorder, scars, autoimmune conditions or  
PT angiogenesis.  
XX  
XX Claim 2; SEQ ID NO 1; 260pp; English.  
PS  
XX The invention relates to an isolated antibody or its portion (I) capable  
CC of specifically binding to an epitope of a heparanase protein. Also  
CC described: (1) a cell line (II) for producing a monoclonal antibody or  
CC its portion, comprising a cell line for producing (I); (2) a  
CC pharmaceutical composition comprising (I) and a carrier; and (3) an  
CC affinity medium (III) for binding human heparanase polypeptides,  
CC comprising (I) immobilized to a chemically inert, insoluble carrier. (I)  
CC useful for treating a subject suffering from a pathological condition,  
CC which involves administering (I) to the subject. (I) is useful for  
CC preventing and treating heparanase-related disorder or condition chosen  
CC from inflammatory disorder, wound, scar, vasculopathy, autoimmune  
CC condition, angiogenesis, cell proliferation, cancerous condition, tumor  
CC cell proliferation, invasion of circulating tumor cells and metastatic  
CC disease. (I) is useful for detecting the presence of heparanase  
CC polypeptide in a sample. (I) is useful for detecting heparanase-related  
CC disease or condition in a subject such as vertebrate, preferably mammal  
CC e.g., human. The heparanase-related disorder or condition further  
CC includes renal disease or disorder chosen from diabetic nephropathy,  
CC glomerulosclerosis, nephrotic syndrome, minimal change nephrotic syndrome  
CC and renal cell carcinoma. The present sequence represents the 45 kDa  
CC subunit of the human mature processed heparanase dimer, which is used in  
CC the exemplification of the present invention.  
XX  
XX Sequence 386 AA;  
SQ  
Query Match 100.0%; Score 2020; DB 9; Length 386;  
Best Local Similarity 100.0%; Pred. No. 4.9e-202;  
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKFKNSTYSSRSDVLYTFANCSSGLDLIFGLNALLRTADLQWSSNAQLLDYCSSKGYN 60

Db 1 KKFKNSTYSRSSVDVLYTFANCSEGLDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYN 60  
QY 61 ISWELGNEPNSFLKKADIFINGSLQGEDFIQLHKLKSTFKNAKLYGPDVGGQPRRTAK 120  
Db 61 ISWELGNEPNSFLKKADIFINGSLQGEDFIQLHKLKSTFKNAKLYGPDVGGQPRRTAK 120  
QY 121 MKLSFLKAGGEVDSVTHHYYLNGRTATREDFLNPDVLDIFISSVKQVQVVESTPRGK 180  
Db 121 MKLSFLKAGGEVDSVTHHYYLNGRTATREDFLNPDVLDIFISSVKQVQVVESTPRGK 180  
QY 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240  
Db 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240  
QY 241 FDPDPYWLSSLFLFKLVGTGKVMASVQGSKRRKLRYVYLHCTNTDNPRYKEGDLTLYAINL 300  
Db 241 FDPDPYWLSSLFLFKLVGTGKVMASVQGSKRRKLRYVYLHCTNTDNPRYKEGDLTLYAINL 300  
QY 301 HNVTKYLRLPYPPFSNKQVDKYLRLPLGPHGLLSKSVQNLGLTLKMWDDQTLPLMEKPLR 360  
Db 301 HNVTKYLRLPYPPFSNKQVDKYLRLPLGPHGLLSKSVQNLGLTLKMWDDQTLPLMEKPLR 360  
QY 361 PGSSGLGPAFYSYFFVIRNAKVAACI 386  
Db 361 PGSSGLGPAFYSYFFVIRNAKVAACI 386

RESULT 6  
ADY27061  
ID ADY27061 standard; protein; 460 AA.  
XX  
AC ADY27061;  
XX  
DT 05-MAY-2005 (first entry)  
XX  
DE Heparanase inhibitor protein #4.  
XX  
KW Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;  
KW neurological disease; viral infection; infection; cytostatic;  
KW antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;  
KW heparanase modulator; enzyme purification.  
XX  
OS Homo sapiens.  
XX  
PN WO2005016227-A2.  
XX  
PD 24-FEB-2005.  
XX  
PF 12-AUG-2004; 2004WO-11000744.  
XX  
PR 14-AUG-2003; 2003US-0494800P.  
PR 12-JAN-2004; 2004US-0535492P.  
XX  
PA (INSI-) INSIGHT BIOPHARMACEUTICALS LTD.  
XX  
PI Van-Gelder JM, Miron D;  
XX  
XX WPI; 2005-182203/19.  
DR  
XX  
XX Regulating heparanase activity, useful for treating heparanase-associated  
PT diseases (e.g. cancer, inflammation, cardiovascular diseases,  
PT neurological diseases or viral diseases) comprises modulating heparanase  
PT activation.  
XX  
PS Disclosure; SEQ ID NO 37; 211pp; English.  
XX  
XX The invention relates to a method of regulating heparanase activity in a  
CC tissue or regulating a biological process depending at least in part on  
CC heparanase activity comprising modulating heparanase activation. The  
CC invention also relates to methods of treating a heparanase- or heparin  
CC binding protein-associated disease or disorder in a subject, a  
CC pharmaceutical composition for use in the treatment of a heparanase-

CC associated disease or disorder comprising a therapeutic amount of an  
CC agent capable of modulating heparanase activation and a pharmaceutical  
CC carrier or diluent, a method of identifying a protease activator of  
CC heparanase, a protease substrate mimetic comprising a peptide  
CC representing a subset or all substrate residues or cleavage sites of  
CC human heparanase or an equivalent non-human heparanase, a method of  
CC producing active heparanase and a method of modulating an adhesion  
CC activity of heparanase. The composition and methods are useful for  
CC modulating heparanase activation and for treating heparanase-associated  
CC diseases or disorders such as cancer, inflammation, cardiovascular  
CC diseases, neurological diseases or viral infections. This sequence  
CC represents a heparanase inhibitor protein used in the scope of the  
CC invention.  
XX  
SQ Sequence 460 AA;  
Query Match 100.0%; Score 2020; DB 9; Length 460;  
Best Local Similarity 100.0%; Pred. No. 6.4e-202;  
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKFKNSTYSRSSVDVLYTFANCSEGLDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYN 60  
Db 75 KKFKNSTYSRSSVDVLYTFANCSEGLDLIFGLNALLRTADLQWNSNAQLLDYCSSKGYN 134  
QY 61 ISWELGNEPNSFLKKADIFINGSLQGEDFIQLHKLKSTFKNAKLYGPDVGGQPRRTAK 120  
Db 135 ISWELGNEPNSFLKKADIFINGSLQGEDFIQLHKLKSTFKNAKLYGPDVGGQPRRTAK 194  
QY 121 MKLSFLKAGGEVDSVTHHYYLNGRTATREDFLNPDVLDIFISSVKQVQVVESTPRGK 180  
Db 195 MKLSFLKAGGEVDSVTHHYYLNGRTATREDFLNPDVLDIFISSVKQVQVVESTPRGK 254  
QY 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240  
Db 255 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 314  
QY 241 FDPDPYWLSSLFLFKLVGTGKVMASVQGSKRRKLRYVYLHCTNTDNPRYKEGDLTLYAINL 300  
Db 315 FDPDPYWLSSLFLFKLVGTGKVMASVQGSKRRKLRYVYLHCTNTDNPRYKEGDLTLYAINL 374  
QY 301 HNVTKYLRLPYPPFSNKQVDKYLRLPLGPHGLLSKSVQNLGLTLKMWDDQTLPLMEKPLR 360  
Db 375 HNVTKYLRLPYPPFSNKQVDKYLRLPLGPHGLLSKSVQNLGLTLKMWDDQTLPLMEKPLR 434  
QY 361 PGSSGLGPAFYSYFFVIRNAKVAACI 386  
Db 435 PGSSGLGPAFYSYFFVIRNAKVAACI 460  
RESULT 7  
ADZ18996  
ID ADZ18996 standard; protein; 492 AA.  
XX  
AC ADZ18996;  
XX  
DT 16-JUN-2005 (first entry)  
XX  
DE Hep106 construct protein.  
XX  
KW Enzyme engineering; heparanase; hep106; metastasis; autoimmune disease;  
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
KW immunosuppressive; enzyme.  
OS Synthetic.  
XX  
PN WO2005030962-A1.  
XX  
PD 07-APR-2005.  
XX  
PF 17-SEP-2004; 2004WO-EP010517.  
XX  
PR 26-SEP-2003; 2003US-0506479P.  
PR 20-JAN-2004; 2004US-0537729P.

```

XX PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
XX PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;
XX PD WPI: 2005-273382/28.
XX DR N-PSDB; ADZ18997.
XX PF Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX PS Example 2; SEQ ID NO 17; 65pp; English.
XX SS The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hepi06 construct protein
CC used in the scope of the invention.
XX SQ Sequence 492 AA;
Query Match 100.0%; Score 2020; DB 9; Length 492;
Best Local Similarity 100.0%; Pred. No. 7.2e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60
DB 107 KKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 166
QY 61 ISWELGNENPSFLKKADIFINGSQLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 120
DB 167 ISWELGNENPSFLKKADIFINGSQLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 226
QY 121 MLKSFLLKAGEVIDSVTWHYYLNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRP GK 180
DB 227 MLKSFLLKAGEVIDSVTWHYYLNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRP GK 286
QY 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
DB 287 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 346
QY 241 FDPPLDYWLSLLFKLVGTVKVLMAVQGSKRRKRLRVYLHCTNTDNPRYKGGDLTYAINL 300
DB 347 FDPPLDYWLSLLFKLVGTVKVLMAVQGSKRRKRLRVYLHCTNTDNPRYKGGDLTYAINL 406
QY 301 HNVTKYLRLPYFPFSNKQVDKYLRLPLGPHGLLSKSVQLNGLTLKMVDDQTLPLMEKPLR 360
DB 407 HNVTKYLRLPYFPFSNKQVDKYLRLPLGPHGLLSKSVQLNGLTLKMVDDQTLPLMEKPLR 466
QY 361 PGSSGLGLPAFSYSFFVIRNAKVAACI 386
DB 467 PGSSGLGLPAFSYSFFVIRNAKVAACI 492
RESULT 8
ID ADZ18999
ID ADZ18999 standard; protein; 495 AA.
XX AC ADZ18999;
XX DT 16-JUN-2005 (first entry)
XX DE Hep109 construct protein.
XX KW Enzyme engineering; heparanase; hepi09; metastasis; autoimmune disease;
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
KW immunosuppressive; enzyme.
XX

```

```

OS Synthetic.
XX WO2005030962-A1.
XX PD 07-APR-2005.
XX PF 17-SEP-2004; 2004WO-EP010517.
XX PR 26-SEP-2003; 2003US-0506479P.
XX PR 20-JAN-2004; 2004US-0537729P.
XX PA (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
XX PI Lahm A, Nardella C, Pallaro M, Steinkuhler C;
XX DR WPI: 2005-273382/28.
XX DR N-PSDB; ADZ18998.
XX SS Synthetic nucleic acid for e.g. inhibitor screening, comprises a
PT nucleotide sequence that encodes mammalian heparanase protein and has two
PT consensus cleavage sites located between specific nucleotide encoding
PT residues.
XX PS Example 2; SEQ ID NO 20; 65pp; English.
XX SS The invention relates to a synthetic nucleic acid molecule that encodes
CC mammalian heparanase protein, where the nucleic acid comprises two
CC consensus cleavage sites recognized by endoproteinase. The sequences are
CC useful for expressing mammalian heparanase in non-mammalian cells and in
CC inhibitor screening assays for the development of therapeutics or
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
CC and/or inflammation. This sequence represents a hepi09 construct protein
CC used in the scope of the invention.
XX SQ Sequence 495 AA;
Query Match 100.0%; Score 2020; DB 9; Length 495;
Best Local Similarity 100.0%; Pred. No. 7.2e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60
DB 110 KKFKNSTYSRSSVDVLYTFANCSGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 169
QY 61 ISWELGNENPSFLKKADIFINGSQLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 120
DB 170 ISWELGNENPSFLKKADIFINGSQLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 229
QY 121 MLKSFLLKAGEVIDSVTWHYYLNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRP GK 180
DB 230 MLKSFLLKAGEVIDSVTWHYYLNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRP GK 289
QY 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
DB 290 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 349
QY 241 FDPPLDYWLSLLFKLVGTVKVLMAVQGSKRRKRLRVYLHCTNTDNPRYKGGDLTYAINL 300
DB 350 FDPPLDYWLSLLFKLVGTVKVLMAVQGSKRRKRLRVYLHCTNTDNPRYKGGDLTYAINL 409
QY 301 HNVTKYLRLPYFPFSNKQVDKYLRLPLGPHGLLSKSVQLNGLTLKMVDDQTLPLMEKPLR 360
DB 410 HNVTKYLRLPYFPFSNKQVDKYLRLPLGPHGLLSKSVQLNGLTLKMVDDQTLPLMEKPLR 469
QY 361 PGSSGLGLPAFSYSFFVIRNAKVAACI 386
DB 470 PGSSGLGLPAFSYSFFVIRNAKVAACI 495
RESULT 9
ADZ19000
ID ADZ19000 standard; protein; 501 AA.
XX

```



```
AC ADZ19000;
XX
XX 16-JUN-2005 (first entry)
XX DE
XX HepG3 construct protein.
XX
XX Enzyme engineering; heparanase; hepgs3; metastasis; autoimmune disease;
XX inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
XX immunosuppressive; enzyme.
XX
XX Synthetic.
XX
XX WO2005030962-A1.
XX
XX 07-APR-2005.
XX
XX 17-SEP-2004; 2004WO-EP010517.
XX
XX 26-SEP-2003; 2003US-0506479P.
XX
XX 20-JAN-2004; 2004US-0537729P.
XX
XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
XX
XX Lahm A, Nardella C, Pallaro M, Steinkuhler C;
XX
XX WPI; 2005-273382/28.
XX
XX N-PSDB; ADZ19001.
XX
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
XX nucleotide sequence that encodes mammalian heparanase protein and has two
XX consensus cleavage sites located between specific nucleotide encoding
XX residues.
XX
XX Example 2; SEQ ID NO 21; 65pp; English.
XX
XX The invention relates to a synthetic nucleic acid molecule that encodes
XX mammalian heparanase protein, where the nucleic acid comprises two
XX consensus cleavage sites recognized by endoproteinase. The sequences are
XX useful for expressing mammalian heparanase in non-mammalian cells and in
XX inhibitor screening assays for the development of therapeutics or
XX pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
XX and/or inflammation. This sequence represents a hepgs3 construct protein
XX used in the scope of the invention.
XX
XX Sequence 501 AA;
Query Match 100.0%; Score 2020; DB 9; Length 501;
Best Local Similarity 100.0%; Pred. No. 7.4e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKFKNSTYSSSDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYN 60
DB 116 KKFKNSTYSSSDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYN 175
QY 61 ISWELGNEPNSFLKKADIFINGSQGLDFTQLHKLLRSTFKNAKLYGPDVQGPRTAK 120
DB 176 ISWELGNEPNSFLKKADIFINGSQGLDFTQLHKLLRSTFKNAKLYGPDVQGPRTAK 235
QY 121 MLKSFUKAGEVIDSVTWHYYLNGRTATREDFLNPDVLDIFISSVQKVQVVESTRPGK 180
DB 236 MLKSFUKAGEVIDSVTWHYYLNGRTATREDFLNPDVLDIFISSVQKVQVVESTRPGK 295
QY 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
DB 296 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 355
QY 241 FDPPLPWLSLLPKLVGTQVLMASVQGSRRKRLRVYLHCTNTDNPYKGGDLTYAINDL 300
DB 356 FDPPLPWLSLLPKLVGTQVLMASVQGSRRKRLRVYLHCTNTDNPYKGGDLTYAINDL 415
QY 301 HNYTKYLRPLYPFNSKQVDYLLRPLGPHGLLSKSVQNLGLTLKQVDDQTLPLPLMEKPLR 360
DB 416 HNYTKYLRPLYPFNSKQVDYLLRPLGPHGLLSKSVQNLGLTLKQVDDQTLPLPLMEKPLR 475
```

```
QY 361 PGSSGLGLPAFSYSPFVIRNAKVAACI 386
DB 476 PGSSGLGLPAFSYSPFVIRNAKVAACI 501
RESULT 10
ADZ19005
ID ADZ19005 standard; protein; 507 AA.
XX
XX AC ADZ19005;
XX
XX 16-JUN-2005 (first entry)
XX
XX HepG6 construct protein.
XX
XX Enzyme engineering; heparanase; hepgs6; metastasis; autoimmune disease;
XX inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;
XX immunosuppressive; enzyme.
XX
XX Synthetic.
XX
XX WO2005030962-A1.
XX
XX 07-APR-2005.
XX
XX 17-SEP-2004; 2004WO-EP010517.
XX
XX 26-SEP-2003; 2003US-0506479P.
XX
XX 20-JAN-2004; 2004US-0537729P.
XX
XX (RICE-) IST RICERCHIE BIOL MOLECOLARE ANGELETTI.
XX
XX Lahm A, Nardella C, Pallaro M, Steinkuhler C;
XX
XX WPI; 2005-273382/28.
XX
XX N-PSDB; ADZ19003.
XX
XX Synthetic nucleic acid for e.g. inhibitor screening, comprises a
XX nucleotide sequence that encodes mammalian heparanase protein and has two
XX consensus cleavage sites located between specific nucleotide encoding
XX residues.
XX
XX Example 2; SEQ ID NO 26; 65pp; English.
XX
XX The invention relates to a synthetic nucleic acid molecule that encodes
XX mammalian heparanase protein, where the nucleic acid comprises two
XX consensus cleavage sites recognized by endoproteinase. The sequences are
XX useful for expressing mammalian heparanase in non-mammalian cells and in
XX inhibitor screening assays for the development of therapeutics or
XX pharmaceuticals for inhibiting or treating metastasis, autoimmune disease
XX and/or inflammation. This sequence represents a hepgs6 construct protein
XX used in the scope of the invention.
XX
XX Sequence 507 AA;
Query Match 100.0%; Score 2020; DB 9; Length 507;
Best Local Similarity 100.0%; Pred. No. 7.5e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKFKNSTYSSSDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYN 60
DB 122 KKFKNSTYSSSDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYN 181
QY 61 ISWELGNEPNSFLKKADIFINGSQGLDFTQLHKLLRSTFKNAKLYGPDVQGPRTAK 120
DB 182 ISWELGNEPNSFLKKADIFINGSQGLDFTQLHKLLRSTFKNAKLYGPDVQGPRTAK 241
QY 121 MLKSFUKAGEVIDSVTWHYYLNGRTATREDFLNPDVLDIFISSVQKVQVVESTRPGK 180
DB 242 MLKSFUKAGEVIDSVTWHYYLNGRTATREDFLNPDVLDIFISSVQKVQVVESTRPGK 301
QY 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 240
```

Db	302	KWVLEGTSSAYCGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN	361
Qy	241	FDPLPDYWSLLFKKLVTGTVLMASVQGSKRRKLRVYLHCTNTDNPRIYEGDILTYAINL	300
Db	362	FDPLPDYWSLLFKKLVTGTVLMASVQGSKRRKLRVYLHCTNTDNPRIYEGDILTYAINL	421
Qy	301	HNVTKYLRPLPYFPFSNKQVDKYLLRPLGPHGLLSKSVQNLGLTKWVDDQDTLPPLMEKPLR	360
Db	422	HNVTKYLRPLPYFPFSNKQVDKYLLRPLGPHGLLSKSVQNLGLTKWVDDQDTLPPLMEKPLR	481
Qy	361	PGSSIGLPAPFSYSPFVIRNAKVAACI	386
Db	482	PGSSIGLPAPFSYSPFVIRNAKVAACI	507

RESULT 11	
ADY27058	
ID	ADY27058 standard; protein; 508 AA.
XX	
XX	ADY27058;
AC	
XX	
XX	05-MAY-2005 (first entry)
XX	
DE	Human inactive heparanase protein.
XX	
KW	Heparanase; cancer; neoplasm; inflammation; cardiovascular disease;
KW	neurological disease; viral infection; infection; cytostatic;
KW	antiinflammatory; cardiovascular-gen.; neuroprotective; virucide;
KW	protease; enzyme; enzyme purification.
XX	
OS	Homo sapiens.
XX	
PN	WO2005016227-A2.
XX	
XX	24-FEB-2005.
PD	
XX	
XX	12-AUG-2004; 2004WO-IL000744.
PF	
XX	
PR	14-AUG-2003; 2003US-0494800P.
PR	12-JAN-2004; 2004US-0535492P.
XX	
XX	(INSI-) INSIGHT BIOPHARMACEUTICALS LTD.
PA	
XX	
PI	Van-Gelder JM, Miron D;
XX	
XX	WPI; 2005-182203/19.
DR	
XX	
PT	Regulating heparanase activity, useful for treating heparanase-associated
PT	diseases (e.g. cancer, inflammation, cardiovascular diseases,
PT	neurological diseases or viral diseases) comprises modulating heparanase
PT	activation.
XX	
PS	Claim 257; SEQ ID NO 34; 211pp; English.

The invention relates to a method of regulating heparanase activity in a tissue or regulating a biological process depending at least in part on heparanase activity comprising modulating heparanase activation. The heparanase activity comprises modulating heparanase activation. The invention also relates to methods of treating a heparanase- or heparin binding protein-associated disease or disorder in a subject, a pharmaceutical composition for use in the treatment of a heparanase-associated disease or disorder comprising a therapeutic amount of an agent capable of modulating heparanase activation and a pharmaceutical carrier or diluent, a method of identifying a protease activator of heparanase, a protease substrate mimetic comprising a peptide representing a subset or all substrate residues or cleavage sites of human heparanase or an equivalent non-human heparanase, a method of producing active heparanase and a method of modulating an adhesion activity of heparanase. The composition and methods are useful for modulating heparanase activation and for treating heparanase-associated diseases or disorders such as cancer, inflammation, cardiovascular diseases, neurological diseases or viral infections. This sequence represents a human inactive heparanase protein used in the scope of the

CC	invention.
XX	
SQ	Sequence 508 AA;
	Query Match 100.0%; Score 2020; DB 9; Length 508; Best Local Similarity 100.0%; Pred. No. 7.5e-202; Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY	1 KKFKNSTYSRSGSDVLYTFANCSEGLDLIFGLNALLRTADLWNSSNAQLLLDYCSSKGYN 60   DB 123 KKFKNSTYSRSGSDVLYTFANCSEGLDLIFGLNALLRTADLWNSSNAQLLLDYCSSKGYN 182 
QY	61 ISWELGNEPNSFLKKAIDIFINGSOLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRKTA 120 
DB	183 ISWELGNEPNSFLKKAIDIFINGSOLGEDFIQLHKLRLKSTFKNAKLYGPDVGQPRRKTA 242 
QY	121 MLKSFLKAGGVIVSVTWHYYLNGRATREDFLNPVDLDIFISSVKVFQVVESTRPCK 180 
DB	243 MLKSFLKAGGVIVSVTWHYYLNGRATREDFLNPVDLDIFISSVKVFQVVESTRPCK 302 
QY	181 KVMIGETSSAYGGCGAPILSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 240 
DB	303 KVMIGETSSAYGGCGAPILSDTFAAGFMWLDKLGLSARMGIEVVMRQVFFGAGNYHLVDEN 362 
QY	241 FDPPLDYLWLSLLFKKLVGTKVLMASVOGSKRRKLRVYLHCTNTDNPRYKEGDLTLIAINL 300 
DB	363 FDPPLDYLWLSLLFKKLVGTKVLMASVOGSKRRKLRVYLHCTNTDNPRYKEGDLTLIAINL 422 
QY	301 HNVTKYLRLPYPPFNKQVDKYLLRPLGPHGLLSKSVQLNGLTLMKVVDDQTLPPLMEKPLR 360 
DB	423 HNVTKYLRLPYPPFNKQVDKYLLRPLGPHGLLSKSVQLNGLTLMKVVDDQTLPPLMEKPLR 482 
QY	361 PGSSILGLPAFSYSFFVIRNAKVAACI 386 
DB	483 PGSSILGLPAFSYSFFVIRNAKVAACI 508 
RESULT 12	
ADZ19006	
ID	ADZ19006 standard; protein; 526 AA.
XX	
AC	ADZ19006;
XX	
DT	16-JUN-2005 (first entry)
XX	
DE	Hephaluro construct protein.
XX	
KW	Enzyme engineering; heparanase; hepHaluro; metastasis; autoimmune disease; inflammation; neoplasm; immune disorder; anti-inflammatory; cytostatic; immunosuppressive; enzyme.
KW	
XX	
OS	Synthetic.
XX	
PN	WO2005030962-A1.
PD	
PD	07-APR-2005.
XX	
PF	17-SEP-2004; 2004WO-EPO10517.
XX	
XX	26-SEP-2003; 2003US-0506479P.
PR	
PR	20-JAN-2004; 2004US-0537729P.
XX	
XX	(RICE-) IST RICERCH BIOL MOLECOLARE ANGELETTI.
PA	
XX	
PI	Lahn A, Nardella C, Pallaro M, Steinkuhler C;
XX	
DR	WPI; 2005-273382/28.
DR	N-PSDB; ADZ19007.
XX	
PT	Synthetic nucleic acid for e.g. inhibitor screening, comprises a nucleotide sequence that encodes mammalian heparanase protein and has two consensus cleavage sites located between specific nucleotide encoding residues.

XX Example 2; SEQ ID NO 27; 65pp; English.

PS The invention relates to a synthetic nucleic acid molecule that encodes

XX mammalian heparanase protein, where the nucleic acid comprises two

CC consensus cleavage sites recognized by endoproteinase. The sequences are

CC useful for expressing mammalian heparanase in non-mammalian cells and in

CC inhibitor screening assays for the development of therapeutics or

CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease

CC and/or inflammation. This sequence represents a hephyaluro connective

CC protein used in the scope of the invention.

XX Sequence 526 AA;

SQ

Query Match 100.0%; Score 2020; DB 9; Length 526;

Best Local Similarity 100.0%; Pred. No. 8e-202;

Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYN 60

DB 141 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYN 200

QY 61 ISWELGNEPNSFLKKADIFINGSQGLGDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 120

DB 201 ISWELGNEPNSFLKKADIFINGSQGLGDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 260

QY 121 MLKSFLLKAGGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTPRGK 180

DB 261 MLKSFLLKAGGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTPRGK 320

QY 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240

DB 321 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 380

QY 241 FDPDPYWLSSLFVKLVGTVMASVQGSRRKRLRVYLHCTNTDNPYKEGDLTLYAINL 300

DB 381 FDPDPYWLSSLFVKLVGTVMASVQGSRRKRLRVYLHCTNTDNPYKEGDLTLYAINL 440

QY 301 HNVTKYLRPLYPFNSKQVDKYLRPLGPHGLLSKSVQNLGLTLKWVDDQTLPLMEKPLR 360

DB 441 HNVTKYLRPLYPFNSKQVDKYLRPLGPHGLLSKSVQNLGLTLKWVDDQTLPLMEKPLR 500

QY 361 PGSSGLGPAFSYFFVIRNAKVAACI 386

DB 501 PGSSGLGPAFSYFFVIRNAKVAACI 526

RESULT 13

ID ABB07815

AC ABB07815 standard; protein; 527 AA.

XX ABB07815;

XX 03-JUL-2002 (first entry)

DT

XX Chicken signal peptide/human heparanase chimeric protein sequence.

DE Heparanase; catalytic; cytostatic; antiviral; antibacterial; enzyme;

XX anti-protozoan; neuroprotective; heparin; chicken; human; chimeric.

KW

XX Synthetic.

OS Gallus gallus.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..19

FT /note= "chicken heparanase signal peptide"

FT Protein 20..527

FT /note= "human heparanase mature protein"

XX US2002034810-A1.

XX 21-MAR-2002.

XX 16-AUG-2001; 2001US-00930218.

XX 20-SEP-2000; 2000US-00666390.

XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX Goldshmidt O, Pecker I, Vlodavsky I, Michal I, Zcharia E;

XX WPI; 2002-338926/37.

XX N-PSDB; ABL40753.

XX Nucleic acid encoding avian and reptile heparanase polypeptide is useful

XX to treat various heparin-related disorders and the signal peptide is

XX useful in production of membrane-targeted or secreted recombinant

XX proteins.

XX Disclosure; Page 26-28; 39pp; English.

XX The invention relates to an isolated avian and reptile nucleic acid,

XX encoding a polypeptide with heparanase catalytic activity. The signal

XX peptide of the nucleic acid can be used to express membrane-associated or

XX secreted proteins in heterologous expression systems. The encoded

XX polypeptides can be used to prevent tumour angiogenesis, metastasis and

XX invasion, and to intervene with pathologies associated with impaired

XX heparin-binding growth factors, cellular responses to heparin-binding

XX growth factors and cytokines, cell interaction with plasma lipoproteins,

XX cellular susceptibility to viral, protozoa and bacterial infections or

XX disintegration of neurodegenerative plaques. The present sequence

XX represents a chicken signal peptide/human heparanase chimeric protein

XX sequence

SQ Sequence 527 AA;

Query Match 100.0%; Score 2020; DB 5; Length 527;

Best Local Similarity 100.0%; Pred. No. 8e-202;

Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYN 60

DB 142 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWSSNAQLLLDYCSSKGYN 201

QY 61 ISWELGNEPNSFLKKADIFINGSQGLGDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 120

DB 202 ISWELGNEPNSFLKKADIFINGSQGLGDFIQLHKLRLKSTFKNAKLYGPDVGQPRRTAK 261

QY 121 MLKSFLLKAGGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTPRGK 180

DB 262 MLKSFLLKAGGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTPRGK 321

QY 181 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240

DB 322 KWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 381

QY 241 FDPDPYWLSSLFVKLVGTVMASVQGSRRKRLRVYLHCTNTDNPYKEGDLTLYAINL 300

DB 382 FDPDPYWLSSLFVKLVGTVMASVQGSRRKRLRVYLHCTNTDNPYKEGDLTLYAINL 441

QY 301 HNVTKYLRPLYPFNSKQVDKYLRPLGPHGLLSKSVQNLGLTLKWVDDQTLPLMEKPLR 360

DB 442 HNVTKYLRPLYPFNSKQVDKYLRPLGPHGLLSKSVQNLGLTLKWVDDQTLPLMEKPLR 501

QY 361 PGSSGLGPAFSYFFVIRNAKVAACI 386

DB 502 PGSSGLGPAFSYFFVIRNAKVAACI 527

RESULT 14

ID ABB02018

AC ABB02018 standard; protein; 527 AA.

XX ABB02018;

XX

DT 12-FEB-2004 (first entry)  
XX Chimeric human-chicken heparanase protein.  
XX  
KW Chicken; heparanase; tumour cell metastasis; inflammation; autoimmunity;  
KW wound healing; angiogenesis; restenosis; Genstmann-Straussler Syndrome;  
KW neurodegenerative disease; atherosclerosis; Creutzfeldt-Jakob disease;  
KW infection; Scrapie; Alzheimer's disease; protein therapy; cytostatic;  
KW immunosuppressive; vulnery; bactericide; anti-angiogenic; virucide;  
KW antisclerotic; neuroprotective; protozoacide; chimeric; fusion protein;  
KW enzyme; human.  
XX  
OS Chimeric - Gallus gallus.  
OS Chimeric - Homo sapiens.  
XX  
PN US2003180788-A1.  
XX  
XX 25-SEP-2003.  
XX  
XX 08-MAY-2003; 2003US-00431438.  
XX  
XX 20-SEP-2000; 2000US-00666390.  
XX 16-AUG-2001; 2001US-00930218.  
XX  
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.  
PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.  
XX  
XX Goldshmidt O, Pecker I, Vlodosky I, Michal I, Zcharia E;  
PI WPI; 2003-843931/78.  
DR N-PSDB; AAD63532.  
XX  
XX Recombinant jungle red fowl (Gallus gallus) heparanase protein, useful  
PT for treating cancers, microbial infections and aiding wound healing.  
XX  
XX Example; Page 26-28; Opp; English.  
XX  
XX The present invention relates to novel jungle red fowl heparanase protein  
CC and polynucleotides encoding such proteins. Heparanase sequences can be  
CC used to develop treatments for various diseases, to develop diagnostic  
CC assays for these diseases and to provide new tools for basic and directed  
CC research especially in the fields of medicine and biology. They can be  
CC used to develop new drugs to inhibit tumour cell metastasis, inflammation  
CC and autoimmunity. Recombinant heparanase offers a potential treatment for  
CC wound healing, angiogenesis, restenosis, atherosclerosis, inflammation,  
CC neurodegenerative diseases (e.g. Genstmann-Straussler Syndrome, Scrapie,  
CC Creutzfeldt-Jakob disease and Alzheimer's disease) and certain viral and  
CC some bacterial and protozoa infections. Recombinant heparanase can also  
CC be used to neutralise plasma heparin, as a potential replacement of  
CC protamine. Sequences of the invention are used in protein therapy. The  
CC present sequence is chimeric human-chicken heparanase protein  
XX  
XX Sequence 527 AA;  
SQ  
Query Match 100.0%; Score 2020; DB 7; Length 527;  
Best Local Similarity 100.0%; Pred. No. 8e-202;  
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60  
Db 142 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 201  
QY 61 ISWELGNEPNSFLKKADIFINGSQLGDFIQLHKLKRSFTFKNALYGPDPVGPQRRTAK 120  
Db 202 ISWELGNEPNSFLKKADIFINGSQLGDFIQLHKLKRSFTFKNALYGPDPVGPQRRTAK 261  
QY 121 MLKSFUKAGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGK 180  
Db 262 MLKSFUKAGEVIDSVTHHYLYNGRTATREDFLNPDVLDIFISSVQKVFQVVESTRPGK 321  
QY 181 KVMGETSSAYGGGAPLLSDTFFAAGFPWMLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 240  
Db 322 KVMGETSSAYGGGAPLLSDTFFAAGFPWMLDKGLSARMGIEVVMRQVFFGAGNYHLVDEN 381

QY 241 FDPLPDYWLSSLFPKLVGTQKVLMAVSGSKRRKLRVYLHCTNTNDNPRYKGGDLTYAINL 300  
Db 382 FDPLPDYWLSSLFPKLVGTQKVLMAVSGSKRRKLRVYLHCTNTNDNPRYKGGDLTYAINL 441  
QY 301 HNVTKYLRLPYPPFSNKQVDKYLRLPLGPHGLLSKSVQLNGLTKMVDQDTLPPLMEKPLR 360  
Db 442 HNVTKYLRLPYPPFSNKQVDKYLRLPLGPHGLLSKSVQLNGLTKMVDQDTLPPLMEKPLR 501  
QY 361 PGSSGLGLPAFSYSFFVIRNAKVAACI 386  
Db 502 PGSSGLGLPAFSYSFFVIRNAKVAACI 527

## RESULT 15

ADZ19004  
ID ADZ19004 standard; protein; 527 AA.  
XX  
AC ADZ19004;  
XX  
XX 16-JUN-2005 (first entry)  
XX  
DE HepGS4 construct protein.  
XX  
KW Enzyme engineering; heparanase; hepGS4; metastasis; autoimmune disease;  
KW inflammation; neoplasm; immune disorder; antiinflammatory; cytostatic;  
KW immunosuppressive; enzyme.  
XX  
OS Synthetic.  
XX  
PN WO2005030962-A1.  
XX  
PD 07-APR-2005.  
XX  
PF 17-SEP-2004; 2004WO-EP010517.  
XX  
PR 26-SEP-2003; 2003US-0506479P.  
PR 20-JAN-2004; 2004US-0537729P.  
XX  
XX (RICE-) 1ST RICERCHE BIOL MOLECOLARE ANGELETTI.  
PA  
XX Lahm A, Nardella C, Pallaoro M, Steinkuhler C;  
XX  
XX WPI; 2005-273382/28.  
DR N-PSDB; ADZ19002.  
XX  
PT Synthetic nucleic acid for e.g. inhibitor screening, comprises a  
PT nucleotide sequence that encodes mammalian heparanase protein and has two  
PT consensus cleavage sites located between specific nucleotide encoding  
PT residues.  
XX  
XX Example 2; SEQ ID NO 25; 65pp; English.  
PS  
XX

CC The invention relates to a synthetic nucleic acid molecule that encodes  
CC mammalian heparanase protein, where the nucleic acid comprises two  
CC consensus cleavage sites recognized by endoproteinase. The sequences are  
CC useful for expressing mammalian heparanase in non-mammalian cells and in  
CC inhibitor screening assays for the development of therapeutics or  
CC pharmaceuticals for inhibiting or treating metastasis, autoimmune disease  
CC and/or inflammation. This sequence represents a hepGS4 construct protein  
CC used in the scope of the invention.  
XX  
SQ Sequence 527 AA;

Query Match 100.0%; Score 2020; DB 9; Length 527;  
Best Local Similarity 100.0%; Pred. No. 8e-202;  
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 60  
Db 142 KKFKNSTYSRSSVDVLYTFANCGLDLIFGLNALLRTADLQWNSNAQLLLDYCSSKGYN 201  
QY 61 ISWELGNEPNSFLKKADIFINGSQLGDFIQLHKLKRSFTFKNALYGPDPVGPQRRTAK 120

Db	202	ISWELGNPNSEFLKADIFINGSQLGEDFIQLHKLRLKSTFKNAKLYGPDVGGQPRRTAK	261
Qy	121	MLKSFLKAGGEVIDSVTHHYYLNGRTATREDFLNPDVLDIFISSVQKVQVVESTTRPGK	180
Db	262	MLKSFLKAGGEVIDSVTHHYYLNGRTATREDFLNPDVLDIFISSVQKVQVVESTTRPGK	321
Qy	181	KVWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNTHLVNEN	240
Db	322	KVWLGETSSAYGGGAPLLSDTFAAGFMWLDKGLSARMGIEVVMRQVFFGAGNTHLVNEN	381
Qy	241	FDPLPDVWLSLLFKKLVGTKVLMASVQGSRRKLRVYLHCTNTDNPYKEGDLTYAINL	300
Db	382	FDPLPDVWLSLLFKKLVGTKVLMASVQGSRRKLRVYLHCTNTDNPYKEGDLTYAINL	441
Qy	301	HNVTKYLRLPYPFSPNKOVDKYLRLPLGPHGLSKSVQNLGLTLKMVDDQTLPLMEKPLR	360
Db	442	HNVTKYLRLPYPFSPNKOVDKYLRLPLGPHGLSKSVQNLGLTLKMVDDQTLPLMEKPLR	501
Qy	361	PGSSGLPAPFSYSFFVIRNAKVAACI	386
Db	502	PGSSGLPAPFSYSFFVIRNAKVAACI	527

Search completed: June 5, 2006, 12:09:41  
 Job time : 78.9552 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2006, 12:10:07 ; Search time 12.8514 Seconds  
(without alignments)  
2889.939 Million cell updates/sec

Title: US-10-645-659A-1  
Perfect score: 2020  
Sequence: 1 KFKKSTYSRSSVDVLYTFA.....LPAFSYSPFFVIRNAKVAACI 386

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	789.5	39.1	480	2 JC7506	heparanase protein
2	407	20.1	521	2 T45608	hypothetical prote
3	169.5	8.4	190	2 T01953	hypothetical prote
4	105.5	5.2	411	2 S74760	hypothetical prote
5	104.5	5.2	788	1 S00652	phosphoribosylamin
6	103.5	5.1	500	2 D87541	beta-xyloridase [i
7	101.5	5.0	670	2 T38446	microtubule-associ
8	100	5.0	654	2 T14202	NADH2 dehydrogenas
9	98	4.9	644	2 A97268	methionyl-tRNA syn
10	97	4.8	379	2 A69974	cystathionine gamm
11	96	4.8	557	1 ODNC1	cytochrome-c oxida
12	93	4.7	385	2 S28360	Al1 protein - beet
13	95	4.7	511	2 S61166	probable membrane
14	95	4.7	699	2 F95146	DNA topoisomerase
15	95	4.7	701	2 D98014	DNA topoisomerase
16	95	4.7	817	2 H75035	probable membrane
17	94.5	4.7	596	2 T04506	hypothetical prote
18	94.5	4.7	804	2 G71546	probable DNA gyras
19	94	4.7	805	2 C86525	DNA gyrase subunit
20	94	4.7	805	2 H72098	DNA gyrase, chain
21	93.5	4.6	356	2 F64383	hypothetical prote
22	93.5	4.6	883	2 AE0207	conserved hypotet
23	93	4.6	454	2 T20829	probable serine ca
24	93	4.6	492	2 T43859	uroporphyrinogen I
25	93	4.6	816	2 G71127	hypothetical prote
26	93	4.6	837	1 A31842	endo-1,4-beta-xyla
27	92	4.6	485	2 F64165	hypothetical prote
28	92	4.6	804	2 A81701	DNA gyrase, chain
29	92	4.6	879	2 F81453	DNA-directed DNA p

30	91.5	4.5	649	2 F85682	unknown protein en
31	91.5	4.5	782	2 H90823	probable secreted
32	91.5	4.5	1500	1 UQ1348	carbamoyl-phosphat
33	91	4.5	796	2 D97065	transketolase [imp
34	91	4.5	822	2 F93016	penicillin-binding
35	91	4.5	2013	2 A11489	probable peptidogl
36	90.5	4.5	418	2 H97437	cyclopropane-fatty
37	90.5	4.5	418	2 AC2656	cyclopropane-fatty
38	90.5	4.5	455	2 H84955	UDP-N-acetylmuramo
39	90.5	4.5	500	2 E83882	alpha-L-arabinofur
40	90	4.5	604	2 E75119	nitrite reductase
41	90	4.5	847	2 AG1001	phosphoribosylamin
42	89.5	4.4	419	1 S75867	hypothetical prote
43	89.5	4.4	425	2 C97354	hypothetical prote
44	89.5	4.4	828	2 S56250	probable membrane
45	89	4.4	541	1 ODZJ1	cytochrome-c oxida

ALIGNMENTS

RESULT 1

JC7506  
heparanase protein 2a - human  
C;Species: Homo sapiens (man)  
C;Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 09-Jul-2004  
C;Accession: JC7506  
R;McKenzie, E.; Tyson, K.; Stamps, A.; Smith, P.; Turner, P.; Barry, R.; Hircocock, M.; Pat  
Biochem. Biophys. Res. Commun. 276, 1170-1177, 2000  
A;Title: Cloning and expression profiling of Hpa2, a novel mammalian heparanase family me  
A;Reference number: JC7506  
A;Accession: JC7506  
A;Molecule type: mRNA  
A;Residues: 1-480 <MCK>  
A;Cross-references: UNIPROT:Q9HB39; UNIPARC:UPI000003E88A; GB:AF282885  
C;Comment: This protein, an intracellular membrane-bound enzyme, has biological and therape  
therapies.  
C;Genetics:  
A;Gene: hpa2a  
A;Map position: 10q23-10q24  
C;Keywords: heparin binding; membrane bound

Query Match 39.1%; Score 789.5; DB 2; Length 480;  
Best Local Similarity 44.0%; Pred. No. 1e-56;  
Matches 164; Conservative 61; Mismatches 129; Indels 19; Gaps 5;

QY	30	GLN-ALLRTADLQWSSNAQLLDYCSKG-----YNISWELGNEPNSFLKKADIFING	82
DB	107	GLSPAFURFGGKRTDFLQFNLRNPAKSRGGPGPDYLLKNYE--DEPNRYRTHMGRVANG	164
QY	83	SQLGEDPIQLHLRK-STFKNAKLYGPDVGQPRRKAKMLKSLFKAGGEVIDSVTHHY	141
DB	165	SQLGKDYIQLKSLLOPIRYSRASLYGPNRKNVTALLDGFVKVAGSTVDATVQHC	224
QY	142	YLNGRTATREDFLNPVDLIFISSVQVKVQVVESTRPKKVMWLGTSBAYGGAPLLSST	201
DB	225	YIDGRVVKWDFLKRLLDLSLQIRKIQKVNTVTPGKIKWLEGVWVTTTSAGGTNNLSDS	284
QY	202	FAAGFMWDLKGLSARMGIEVWVQVFFGAGNYHLVDENFDPLPDYWLSSLLEPKLVGTVK	261
DB	285	YAAAGFLWLTGLMLANQIDIVIRHSPFDHGNYHLVDQNFNPLPDYWLSSLLYKRLIGPKV	344
QY	262	LMASVQGSKKR-----KLRVYLCTNTDNPYKEGDLTVAINLHNVTKYLRPLPY	312
DB	345	LAVHVAGLQKRPGRVIRDKLRIYACTNNHNNHYVRSITFLINLHRSKKIKLAGT	404
QY	313	FSNKQVDYLLRPLPGHLLSKSVQLNGLTLKMWDDQTLPLMEKPLPGSGSLGPAFSY	372
DB	405	LRDKLVHQYLLQPYQGSEGLKSKSVQLNGQPLVMVDDGTLPELKPRLRAGRTLVI	464
QY	373	SFFVIRNAKVAAC	385
DB	465	GFFVVKVNNALAC	477

## RESULT 2

T45608  
hypothetical protein F13G24.30 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 09-Jul-2004  
C:Accession: T45608  
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba  
submitted to the Protein Sequence Database, December 1999  
A:Reference number: Z23009  
A:Accession: T45608  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-521 <BEV>  
A:A:Cross-references: UNIPROT:Q9SDM1; UNIPARC:UPI000000A497C; EMBL:AL133421  
A:Experimental source: cultivar Columbia; BAC clone F13G24  
C:Genetics:  
A:Map position: 5  
A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3  
A>Note: F13G24.30

Query Match	20.1%	Score	407;	DB	2;	Length	521;		
Best Local Similarity	32.3%;	Pred.	No. 2.4e-25;						
Matches	137;	Conservative	57;	Mismatches	158;	Indels	72;	Gaps	19;
QY	14	DVLYTFANCGLDLIFGLNALLRTADLQ-----WNSSNAQLLLDYCSSKGYNI--SWEL	65						
DB	117	DELNSFLTATGAVVTFGLNALRGHRKLRKAWGGAWDHINTQDFLNTVTSKGYVIDSWEF	176						
QY	66	GNEPNGLFKKADIFINGSQLGEDFIQLHLKLRKSTFKNAKLYGPDVGQP-----RRKTAK	120						
DB	177	GNELSG--SGVGASVSAELYGKGLIVLKDVKINK-VYKNSWLHRPILVAPGGFTEQOQWYTK	233						
QY	121	MLKSFLLKAGGEVIDSVTWHHYILNGRT--ATREDFLNPDLVDIFISSVKVF-----QVVE	174						
DB	234	LLLEI---SGFSVDDVVTHIYNLGSGNDPALVKKIMDPS---YLSQVSKTFKDVNQTIQ	286						
QY	175	STRPGKKVILGETSSAYGGGAPLLSDTFAAGFWMLDKLGLSARMGIEVWVRQVFFGAGNY	234						
DB	287	EHPGWPASPVWGSEGGAYNSGGRHVSDTFIDSFYWLQDGLMSARHNTKVYCRQTLVAG-GFY	345						
QY	235	HLVDE-NFQDPLPDYWSLLPKKLVGTKVLMASVQSKRKLRYVYLHCTNTDTPRYKEGDL	293						
DB	346	GLLEKTFEPNPDYYSALLWHRLMGKGLTAVQTDGPP--QLRVYAHCSK-----GRAGV	397						
QY	294	TLVAINLHNVTXYL-----RLPVPS--NKQVDKYLRLP--	325						
DB	398	TLILLINLSQSDFTVSVNSGINVVLNAESPKKXSLDLTLRPPSWTQSKASDGYLNREEY	457						
QY	326	-LGPHG--LLSKSVQNLGLTKMWDDQTLPLMEKPLRP-GSSLGLPAPFSYSPFVIRNAK	381						
DB	458	HLTPENGVLRSKTMVLNGSKSLKPTATGDIISL-EPVLRSVNSPLNPLPLSMSSFVLNPF	516						
QY	382	VAAC	385						
DB	517	ASAC	520						

### RESULT 3

T01953  
hypothetical protein T2L5.6 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 09-Jul-2004  
C:Accession: T01953  
R:Geisel, C.; Smith, A.; Le, T.  
submitted to the EMBL Data Library, October 1998  
A:Description: The sequence of A. thaliana T2L5.  
A:Reference number: Z14470  
A:Accession: T01953  
A:Status: translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-190 <GEI>

## RESULT 4

S74760  
hypochemical protein slr1617 - *Synechocystis* sp. (strain PCC 6803)  
C;Species: *Synechocystis* sp.  
C;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C;Accession: S74760  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Rec. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
s.  
A;Reference number: S74322; PMID:97061201; PMID:8905231  
A;Accession: S74760  
A;Status: preliminary  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-411 <KAN>  
A;Cross-references: UNIPARC:UPI00000C0C3B; EMBL:D90901; GB:AB001339; NID  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match	5.2%; Score 105.5; DB 2; Length 411;
Best Local Similarity	21.4%; Pred. No. 0.85;
Matches	66; Conservative 54; Mismatches 95; Indels 93; Gaps 17;

  

Qy	2	KFKNSYSRSSVDVLYTFANCSGDLILF---GLNALLRTADLWNSSNAQL-----	49
Db	151	EFLRLPTREQIDI---FAGSTKLDSLAEENIDCIIVHLANPRVYTSNVANGOTLTMLRN	207
Qy	50	LLEDYCSKSG---YNISWEL-----GNEPNSFLKK-----ADIFNGSQLGE	87
Db	208	VIVIDCLAKOITPLIYPSSESIYGAVGTIHADESTPALPRGPYGETKYLAELI---	260
Qy	88	DFIQHLKLKRSTFNQAKLYGPDVGQPRRKTAQMKLFKAGGEVIDSVTHHYYLNCRG	147
Db	261	DHCRTGTLRCALIRSSPVYGSMDKP-----KFIFNFKKASOGQKIYT--HHYING--	311
Qy	148	ATREDFFLPNPV----LDIFISSVOKVQFVVESTRPGKKVMILGETSSAYCGGAPLLSDTFA	203
Db	312	-----NPKLDDLHIIDLSSIVATL-----KSRFIGNLNI-----GTGQSLSLTK	351
Qy	204	AGFMWLDKLGISA-----RMGIEVVMQVPFGAGNYHLVDENFDPLPYWLUSLLFPKLVG	258
Db	352	IAEMIRDLCSSSMIOOEIVNTVEASTAMNYGRAN-HVLID-----WPVIIFE-QG	400

A;Cross-references: UNIPROT:O82604; UNIPARC:UPI0000A8F7D; EMBL:AF096371; NID:g3695386; I  
A;Experimental source: cultivar Columbia

A;Map position: 4  
A;Introns: 36/2; 69/3  
A;Note: T2L5.6  
C;Superfamily: Arabid

	Query Match	8.4%;	Score 169.5;	DB 2;	Length 190;
	Best Local Similarity	27.8%;	Pred. No. 1.6e-06;		
	Matches 54;	Conservative 34;	Mismatches 57;	Indels 49;	Gaps 9;
Qy	225	RQVFFGAGNYHLVD--ENFOPLPDYWLSLFLPKVLVGTKVLMAVSOGSKRRKRLRVYLHCTNT	283		
Db	12	RQSLIG-GNYGLLNTTFTFPDYYYSALLTWQLMGRKALFTTFSGTK--KIRSYTHCA--	66		
Qy	284	DNPYKEGBDLTYAIVNLHV-----TKYLRUPYPFSNKQVDKYLLRPL	326		
Db	67	---RQSKG-ITVLLMLNDLTTTTVAAKVELNNSPSLRHTKHMK-----SYKRASSQLFG--	115		
Qy	327	GPHGLL-----SKSVOLNGTLTKMWDDOTLPPLMEKPLPGSSIGH;PAPS	371		
Db	116	GPNGVIQREYHYLTAKDGNLHSQTMLINGNALQVNSMGDLPPIEPIHINSTEPTIAPYS	175		
Qy	372	YSFFVIRNAKVAAC	385		
Db	176	IVFVHMNVVVVPAC	189		

## RESULT 4

S74760  
hypochemical protein slr1617 - *Synechocystis* sp. (strain PCC 6803)  
C;Species: *Synechocystis* sp.  
C;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C;Accession: S74760  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Rec. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
s.  
A;Reference number: S74322; PMID:97061201; PMID:8905231  
A;Accession: S74760  
A;Status: preliminary  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-411 <KAN>  
A;Cross-references: UNIPARC:UPI00000C0C3B; EMBL:D90901; GB:AB001339; NID  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match	5.2%; Score 105.5; DB 2; Length 411;
Best Local Similarity	21.4%; Pred. No. 0.85;
Matches	66; Conservative 54; Mismatches 95; Indels 93; Gaps 17;

  

Qy	2	KFKNSYSRSSVDVLYTFANCSGLDLIF--GLNALLRTADLOWNSSNAQL-----	49
Db	151	EFLRLPTREQIDI---FAGSTKLDLASSEENIDCIIVHLANPRVYTSNVANGOTLTMLRN	207
Qy	50	LLEDYCSKG---YNISWEL-----GNEPNSFLKK-----ADIFNGSQLGE	87
Db	208	VIVIDCLAKOPLIYPSPSSEIYGAVGTIHADESTPALPRGPYGETKYLAELI---	260
Qy	88	DFIQHLKLKRSTFNQAKLYGPDVGQPRRTAKMKLFKAGGEVIDSVTHHHYYLNCRG	147
Db	261	DHCRTGTLRCALIRSSPVYGSMDKP-----KFIFNFKKASOGQKIYT--HHYING--	311
Qy	148	ATREDFFLPNPV----LDIFISSVOKVQFVVESTRPGKKVMILGETSSAYCGGAPLLSDTFA	203
Db	312	-----NPKXLDLHIDDLISSIVATL-----KSRFIGNLNI-----GTGQSLSLTK	351
Qy	204	AGFMWLDKLGISA-----RMGIEVVMQVPFGAGNYHLVDENFDPLPYWLUSLLFPKKLVG	258
Db	352	IAEMIRDLCSSSMIOOEIVNTVEASTAMNYGRAN-HVLDD-----WPVIIFE-QG	400



D87541  
beta-xylosidase [imported] - Caulobacter crescentus  
C/Species: Caulobacter crescentus  
C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
C/Accession: D87541  
R/Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.;  
N. J.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon-  
B. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A/Title: Complete Genome Sequence of *Caulobacter crescentus*.  
A/Reference number: A87249; MUID:21173698; PMID:11259647  
A/Accession: D87541  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-500 <STO>  
A/Cross-references: UNIPROT:Q9A5U0; UNIPARC:UPI00000C76D1; GB:AE005673; NID:g13423886; PJ2  
C/Genetics:  
A/Gene: CC2357

Query Match 5.1%; Score 103.5; DB 2; Length 500;  
Best Local Similarity 25.2%; Pred. No. 1.6;  
Matches 55; Conservative 36; Mismatches 80; Indels 47; Gaps 15;

QY 8 YSRSSVDVLTFFANCSGLDLIFGLN---ALRTAD---LOW--NSSNAQL-----LLD-- 52  
DB 81 YDMTKIDQLYDALAKGIKPFIELGFTPEANKTSDQTIFYMKGNTSHPKLGPWRDLIDAF 140  
QY 53 -YCSSKGYNI-----SW--ELGNEPN--SFLKKADIFINGSQLGEDFIOLHKLKRSTFKN 103  
DB 141 VHLRLARYGVEEVRTWTFEVEWNEPNLDFWEKAD-----QAAYFELYDV---TARA 188  
QY 104 AKLYGPD--VGQPRRKTKAMKSLF---KAGEVIDSVTHHYHLNG-----RTATREDFL 154  
DB 189 IKATPSLRVGGPATAGAAWPEFLAHVKKSGSAVDFVTHTYGVGGFLDEKGVQDTKL 248  
QY 155 NPDVLDIFISSVKVQFQVVE--STRPGKKVWLGETSSAY 191  
DB 249 SPSP-DAVGVGVRRVREQIEASAFPGPLPLYFTWSTSY 285

RESULT 7  
T38446  
microtubule-associated protein sm4 - fission yeast (*Schizosaccharomyces pombe*)  
C/Species: *Schizosaccharomyces pombe*  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: T38446; T00012  
R/McDougall, R.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL data library, December 1997  
A/Reference number: Z21793  
A/Accession: T38446  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-670 <MCD>  
A/Cross-references: UNIPROT:O42667; UNIPARC:UPI0000135FDD; EMBL:AL009227; PIDN:CAA15832.1  
A/Experimental source: strain 972h-; cosmid c27D7  
R/Yamashita, A.; Watanabe, Y.; Yamamoto, M.  
Genes to Cells 2, 155-166, 1997  
A/Title: Microtubule-associated coiled-coil protein Sm4 is involved in the meiotic devel  
A/Reference number: Z14042; MUID:97311255; PMID:9167972  
A/Accession: T00012  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-670 <YAM>  
A/Cross-references: UNIPARC:UPI0000135FDD; EMBL:AB000269; NID:g3341860; PIDN:BAA31857.1;  
C/Genetics:  
A/Gene: sm4; SPAC27D7.13c  
A/Map position: 1

Query Match 5.0%; Score 101.5; DB 2; Length 670;  
Best Local Similarity 23.5%; Pred. No. 3.6;  
Matches 72; Conservative 42; Mismatches 109; Indels 83; Gaps 16;

QY 6 STYSRSRSDVLYTFAN-----CSGLDLIFGLNALLRTADLQWNSNAQLLDYCS----- 55

```
Db 148 STEELSSFT--TLNLSDTSKUSGLD-----DSSFMEEEFVWVDN--VLQCEKXKFTF 197
QY 56 -SKGYNISWELGNEPNSFLKKA---DIFINGSOLGEDFIQLHKLRLK---STFFNAKLY 107
Db 198 HSKGSYLKENLASE-----LRKGRDLMLCENTALKKIDKLNKELEKVEPQLTFLRSK-- 251
QY 108 GPDVGQPRR-KTAKMLKSPFKAGGEV-----IDSVTHHHYLLNGRATATREDFL 154
Db 252 -NSIEKPRNFRREKFLKFLKELAMQKIKYLKRKLQIRKIPNYKYSDRSLSNKTTPKQDNW 310
QY 155 NPDVLD---IFLSSVQKVFQVVESTPGKKWVLGETSSAYGGAPILLSDTFAAGFWMLDK 211
Db 311 TTQVTPSSLLGVSEVSKVLQL-----KQVQVDITE-----LVKIPK 346
QY 212 LGLSARMGTEVVMQVFFGAGN---YHLVDENFDPLPDYWLSSLFKKLGVTKVLMAVQ 267
Db 347 NPFSEKLITSNVRYNLNIVPGSLDLOFSLTNEF-----VHNSVTYQELLNLKSNNSVD 402
QY 268 GSKRRK 273
Db 403 GVKTRR 408

RESULT 8
TL14202
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Rhizopus stolonifer mitochondrion
C:Species: mitochondrion Rhizopus stolonifer
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: TL14202
R:Paquin, B.; Roewer, I.; Wang, Z.; Lang, B.
submitted to the EMBL Data Library, November 1994
A:Description: A robust fungal phylogeny using the mitochondrially encoded nad5 protein
A:Reference number: Z17884
A:Accession: TL14202
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-654 <PAQ>
A:Cross-references: UNIPROT:P50367; UNIPARC:UPI000013081A; EMBL:U17011; NID:g604923; PID
A:Experimental source: DAOM 148428
C:Genetics:
A:Genome: mitochondrion
A:Note: nad5
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 5.0%; Score 100; DB 2; Length 654;
Best Local Similarity 21.5%; Pred. No. 4.6;
Matches 82; Conservative 60; Mismatches 139; Indels 100; Gaps 17;

QY 30 GLNALILRTADLQWSSNAQLLDYCCKGYNISWELGNEPNSFLKADIFINGSOLGEDF 89
Db 256 GYVLLRSPILFEGSTALILITWVGALTAFPAATTTGLLQND-LKRVIAYSTCSQLGLLF 314
QY 90 -----IQLHKLRLKSTFK-----NAKLYGPDVGQPRRKTAKMLK-----SFLK 127
Db 315 LVCGLSQYNVALFHLVNHAWFALLFELSAGSVIHANDEQDLRKFGSLRLPFTYSMMV 374
QY 128 AGG-----EVIDSVTHHHYLLNGRATATREDFLNPDVLDIFISSVQKVFQV 172
Db 375 IGSLSLMALPFLTGTGYFKDLIIITELAVGHVSFSGN-----LVYWLASVAAVFTA 422
QY 173 VESTRPGKKWVLGETSSAYGGG-----APLIS-----DTFAAGFWMLDKLGLS 215
Db 423 MYSIRSLVLTFLG-----YPNGPKINYNINHEAPLIMAIPLVVLAVFSIFFGVYTK-DLF 476
QY 216 ARMGTEVVMQVFFGAGNTHLVNDFDPLPDYWLSSLFKKLGVTKVLMAVQSKRRKL 275
Db 477 VGMGTDFFYNNAIFIHPNHSILVDTEFG-LP-----MSMKFLPLGSLGTLGTGV-----LA 525
QY 276 VY-----LHCTNTDNPRYKEGDLTIYAINLHN--VTKYLRPLPFPFSNKOVDKYLLR 324
Db 526 IYWI FDELPNKFI STKLGRIYRFFNQKYFDNIYNNLNLKFLNFCYT-TNKILDRGAIE 584
```

```
QY 325 PLOPHGLL-----SKSVQNLNGL 341
Db 585 LVGPYGLVNVFKSASNKVSGL 605

RESULT 9
A97268
methionyl-tRNA synthetase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: A97268
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Dally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: A97268
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-644 <KUR>
A:Cross-references: UNIPROT:O97EW5; UNIPARC:UPI000013658D; GB:AE001437; PIDN:AAK80932.1;
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2991
C:Superfamily: methionyl-tRNA synthetase, dimer-forming

Query Match 4.9%; Score 98; DB 2; Length 644;
Best Local Similarity 18.4%; Pred. No. 6.6;
Matches 78; Conservative 57; Mismatches 146; Indels 142; Gaps 19;

QY 6 STYSRSVDVLYTFANGSGLDLIFGLNALLRTADLQWSSNAQLLDYCCKGYNIS--- 62
Db 24 NTYTVASDALVRFKRLQGYDAFW-----LTCTDEHGQKIORIAEDKGITPKAYV 73
QY 63 -----WELGN-EPNSFLKADIFINGSOLGEDFIQLHKLRLKSTFFNAKLYG--- 108
Db 74 DEIVAGIKDLWKWNNISYDKFIRTTD-----EEHVRAVQKIVKFKFYDNGDIYKSAY 124
QY 109 -----PDVGQPRRKTAKMLKSF--LKAGGEVIDSVTHHHYLLN 144
Db 125 EGMVYCPESFWTETQLVDGKCPDGPVEKTEEAYFFKMSKYADRLIKYIEHPDPIQ 184
QY 145 GRTATRE---DFLNPDPVLDIFISSVQKVFQVVESTRPGKKWVLGETSSAYGGAPLISDT 201
Db 185 PESRKNEMLNFLKPGQLDLCIS-----RSSFDMGIPITFDE 221
QY 202 FAAGFWMLDKL-GLSARMGTEVVMQV--FGAGNYHLVDEN---FDPLPDYWLSSLF-- 253
Db 222 KHVIYVWIDALSNIYITALGYSDNDELYNKFWPADLHVGKDIIRFHTI--YWPIMLMAL 279
QY 254 -----KKLVCTKVLMAVQSKRRKLK-----VYLHCTNTDNPRYKEGDLTIYAINLH 301
Db 280 DLEPLKQVFGHGWL--VDGGKMSKSGNVDPVVLNINEFGTDPPVY----- 324
QY 302 NVTKYLRPLPFPFSN-----KQVDKYLRPLGPHGLLSKSVQLNGLTLKWDVQDTL 351
Db 325 -----YLLHEIPFGSDGLFNNEIFIKKINSDLANDLG--NLVSRTA---AMIEKFDSGIQ 375
QY 352 PPL 354
Db 376 PPV 378

RESULT 10
A69974
cystathionine gamma-synthase homolog yrhB - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Oct-2004
C:Accession: A69974
R:Kunst, F.; Ogasawara, N.; Mooser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrati, E.
```

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen-  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;  
Koester, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Paro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
teuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpetra, P.; Togmon, A.; Tosato, V.; Uchiyama,  
I.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: A69574

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-379 <KUN>

A:Cross-references: UNIPROT:O05394; UNIPARC:UPI0000060808; GB:Z99117; GB:AL009126; NID:9

A:Experimental source: strain 168

C:Genetics:

A:Gene: yzhB

C:Superfamily: cystathionine gamma-synthase

Query Match 4.8%; Score 97; DB 2; Length 379;  
Best Local Similarity 21.7%; Pred No. 3.7;  
Matches 72; Conservative 54; Mismatches 112; Indels 94; Gaps 16;

QY 73 LKADIFINGSQGED-----FIQLHKLKSTFKNAKLVDPDVGQPRRTAKMLKSLK 127

Db 1 MKKKTLMHGGITGDEKTAGVSPVQV---STYQPKA-QGHTGYEYRTNPTRTALE 56

QY 128 AGEVDSVTWHYLYNGRTATREDFLNPDLVDFISSVQKVQVVESTPRGKKVWLGET 187

Db 57 ALVTELESGEAGYAFSGMAA-----ITAVNMLF-----NSGDHVVLT--T 94

QY 188 SSAYGSGAPLLSTPAAGFWMLDKGL-----SARMGLEVMVR---QVFFGAGNVHLV 237

Db 95 DVIYGG-----TYRMTKVLNRLGIESTFVDTSGREEVEKAIKPNTKAIY-----I 140

QY 238 DENFOLDPLDYLKLLFKLVGTVKLVMSVQSKRRKLRVYLHCTNTDNPYKSGDILTYA 297

Db 141 EPTNPL-----LKITDLTLMADIAKAGVLLIVDNTFNTFYFQOPLTLGA 186

QY 298 -NLNHNVTYKLRYPYF-----SNQVDKYL-----RPLGPHG--LLSKSVQLNG 340

Db 187 DIVLSATKYLGCHSDWGLVVTASKELGEELHFQVNSTGGVLPQDSSLNRKIGITLG 246

QY 341 LTLKVVDDQTLF-----PLMEKPLRPGSS 364

Db 247 LRMEAIQDNARKIASFLENHPAVQTLYTPGSS 278

RESULT 11

ONC1

Cytochrome-c oxidase (EC 1.9.3.1) chain I - *Neurospora crassa* mitochondrion

N:Alternate names: cytochrome a3 polypeptide I; cytochrome a3 polypeptide I

C:Species: mitochondrion *Neurospora crassa*

C:Date: 13-Aug-1986 #sequence revision 13-Aug-1986 #text\_change 09-Jul-2004

A:Accession: A04469; S07650; S07651

R:Burgen, G.; Scriven, C.; Machleidt, W.; Werner, S.

EMBO J. 1, 1385-1391, 1982

A:Title: Subunit 1 of cytochrome oxidase from *Neurospora crassa*: nucleotide sequence of

A:Reference number: A00469; MUID:84207889; PMID:6327266

A:Accession: A00469

A:Molecule type: DNA

A:Residues: 1-557 <BUR>

A:Cross-references: UNIPROT:P03945; UNIPARC:UPI0000127F8D; EMBL:X01850; NID:g13119; PIDN

A:Note: the amino end of the mature protein may be 3-Ser

R:Field, D.J.; Somerville, A.; Saville, B.J.; Collins, R.A.

Nucleic Acids Res. 17, 9087-9099, 1989

A:Title: A group II intron in the *Neurospora* mitochondrial coI gene: nucleotide sequence

A:Reference number: S07649; MUID:90067912; PMID:2531370

A:Accession: S07650

A:Molecule type: DNA

A:Residues: 1-73 <PIE>

A:Cross-references: UNIPARC:UPI0000017215B; EMBL:X14669; NID:g13123

R:Collins, R.A.

submitted to the EMBL Data Library, March 1989

A:Reference number: S07651

A:Molecule type: DNA

A:Residues: 1-526, V, 528-540 <COL>

A:Cross-references: UNIPARC:UPI0000017215C; EMBL:X14669; NID:g13123

R:Vassiliev, A.O.; Pleofsky-Vig, N.; Brambl, R.

Proc. Natl. Acad. Sci. U.S.A. 92, 8680-8684, 1995

A:Title: Cytochrome c oxidase in *Neurospora crassa* contains myristic acid covalently link

A:Reference number: A43101; MUID:96004602; PMID:7567996

A:Contents: annotation; modified site

C:Genetics:

A:Gene: COI

A:Genome: mitochondrion

A:Genetic code: SGC3

A:Introns: 72/1; 99/2; 210/3; 249/2

C:Function:

A:Description: the cytochrome-c oxidase complex catalyzes the oxidation of four molecules

ns from the mitochondrial matrix producing two molecules of water and lowering the concn

A:Pathway: oxidative phosphorylation; respiratory chain

A:Note: chain I directly reduces oxygen on the mitochondrial matrix side of the inner-mem

C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology

C:Keywords: chromoprotein; copper; electron transfer; heme; iron; lipoprotein; magnesium;

F:16-462/Domain: cytochrome-c oxidase chain I homology <COI>

F:67,383/Binding site: heme a iron (His) (axial ligands) #status predicted

F:246,295,296/Binding site: copper (His) #status predicted

F:246,295,296/Binding site: copper (His) #status predicted

F:250/Binding site: oxygen (Tyr) #status predicted

F:324/Binding site: myristate (Iys) (covalent) #status experimental

F:373/Binding site: magnesium (His) (shared with chain II) #status predicted

F:381/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 4.8%; Score 96; DB 1; Length 557;

Best Local Similarity 20.9%; Pred. NO. 7.8;

Matches 68; Conservative 43; Mismatches 116; Indels 98; Gaps 16;

QY 99 STFNKALYGPVQCPRRKTAKMLKSLKAGEVIDSVTW-HYYLYNGRTATREDFLNPD 157

Db 265 SAYSNKSVFG-YIG-----MYAMMSIG--ILGFIVSHHMYTVGLDVDTFTAA 313

QY 158 VLDIFISGVQVQVVESTPRGKKVWLGETSAYGGGAPLL-SDTFAAGFMWLDKLG--- 213

Db 314 TLITAVPTGKIFS-----WL---ATCYGGSRLTPSMLFALGFVFMFTIGGLS 359

QY 214 --LSARMGIEVMVRQVFF-----GA-----GNYHLVDE---NFDPL---PDY 247

Db 360 GVVLNANSLDIAFHDTYVVAHFHVLMSGAVFAMFSGWVHWVPKILGLNYNWLKSAQF 419

QY 248 WLSLLFKLVGTVKLVMSVQSKRRKLRVYLHCTNTDNPYKSG----- 291

Db 420 WLLFGVNLTFPPQHFLGLQMPRR-----ISDYPDAPSGWNLSSFGSIVSVWAS 470

QY 292 DLTLYANLHNVTYKLRVLPFPFSNKQVDKYLRLPLGRPLGSLKSVQLNGLTLKVVDDQTL 351

Db 471 WLFYIVYIQLVQGEYAGRPWSIPQFVTDLSL-----ALLNRSY-----PSL 513

QY 352 PPLMEKPLRPGSSIGLPAFYSYFFV 376

Db 514 EWSISSPPKPSFASLPLQSSSFFL 538

RESULT 12

S28360

Al1 protein - beet curly top virus

C:Species: beet curly top virus

C:Date: 07-May-1993 #sequence\_revision 07-May-1993 #text\_change 09-Jul-2004

C:Accession: S28360

R:Stanley, J.; Markham, P.G.; Callis, R.J.; Pinner, M.S.

EMBO J. 5, 1761-1767, 1986  
A:Title: The nucleotide sequence of an infectious clone of the geminivirus beet curly to  
A:Reference number: S28360  
A:Accession: S28360  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-385 <STA>  
A:Cross-references: UNIPROT:Q39485; UNIPARC:UPI00000F979C; GB:M24597; EMBL:X04144; NID:9  
C:Superfamily: tomato golden mosaic virus AL1 protein

Query Match 4.7%; Score 95; DB 2; Length 385;  
Best Local Similarity 21.8%; Pred. No. 5.6;  
Matches 67; Conservative 43; Mismatches 98; Indels 100; Gaps 18;  
Qy 92 LHKLLR---KSTFKNAKLYGPDVGOP-----RRKTAKMLKSLFKAGGEVDSDVT 137  
Db LVLVQLLEKQVITIRLF-DLVSPTRSAHPNIQRAKSSDVKYVDKG---DTIE 139  
Qy 138 WHYYLNGRTA-----TRED-----FLNPDVLD-----IFISSVQKV 169  
Db WGEFQIDGRSARGGQQTANDSYAKALNATSLQALQILKEEQKDYFLQHNLLNNAQKI 199  
Qy 170 FQVVESTPGKKVW-----LGETSSAYGG-----GAPLL-----SDTFA 203  
Db 200 FQ-----RP-PDPWTFPLFLSSLFTNVPEEMQEWADAYFGVDAARPLRYNSIIVEGDSRT 253  
Qy 204 AGFWMLDKLG---LSARMGIEVVMQVFFGAGNYHLVDENPDPLDYWLSSLFKKLVT 259  
Db 254 GKTMMARS LGAHNYITGHLDIFS---PRTYDEVEYNVID---DVPDPTLKMHWKHLIGA 307  
Qy 260 -KVLMSVQGSRRKRLRVYLHCTNTDNP-----RYKEGLTLYAHLNHNVTKYRL 309  
Db 308 QKEWQTNLKYGPRIKVGIPICILLNCPGESSYQOFLEKPENEALKSWTLHNS-TFCKL 366  
Qy 310 PYPFSNKQ 317  
Db 367 QGPLENNQ 374

RESULT 13  
S61166  
probable membrane protein YDR371w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein D9481.7  
C:Species: Saccharomyces cerevisiae  
C:Date: 23-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 09-Jul-2004  
A:Accession: S61166  
R: Ding, H.  
submitted to the EMBL Data Library, June 1995  
A:Description: The sequence of S. cerevisiae cosmid 9481.  
A:Reference number: S61159  
A:Accession: S61166  
A:Molecule type: DNA  
A:Residues: 1-511 <DIN>  
A:Cross-references: UNIPROT:Q06350; UNIPARC:UPI000006A221; EMBL:U28373; NID:9849184; PID  
A:Experimental source: strain S288C (AB972)  
C:Genetics:  
A:Gene: MIPS:YDR371w  
A:Cross-references: SGD:S0002779  
A:Map position: 4R  
C:Superfamily: Serratia marcescens chitinase  
C:Keywords: transmembrane protein  
F:18-34/Domain: transmembrane #status predicted <TMM>

Query Match 4.7%; Score 95; DB 2; Length 511;  
Best Local Similarity 19.0%; Pred. No. 8.4;  
Matches 75; Conservative 70; Mismatches 149; Indels 100; Gaps 17;  
Qy 23 SGLDLIFGLNALLRTRADLQWNSNAQL-----LLDYCSSKGYNTSWELGNE 68  
Db 124 SLENNLYKSLAKNSSELTKSSNNVQNILPGICIGELFYLNKTCSDKKFKVIMSIGW 183  
Qy 69 PMSFLKKADIFANGSLQGDFTQ-----LHKLLRKSTFKNAKLYGPDVGOPR--RKTAKM 121

Db 184 SDS--ENFKIILKDDKLLQLQNFVDSVETMPLRGLDGLDWEFFGNNESEPRGYLKLRM 241  
Qy 122 LKSFKA-GGEVIDSVTHHYLNGRTATREDFLNPDVLDIFISSVQKVQFVWVSTRPGK 180  
Db 242 LRLKLSLESQIFGKRTEDHFQLSIAAPAFK-----KLFYLPITFIDQYVDYNNMT 294  
Qy 181 KWLGETSSAYGGGAPLLST-----PAAGFMWLDKLGSLARMGIEVVMQVFFGAGNYH 235  
Db 295 YDYGSWSETTYGHSNLFSETELNNGNFAMHYM-IDRFQVNSR---KVLGMAAYGR-SFH 349  
Qy 236 LVDENFDPLD--YWLSSLLEFKL-----VGT-----KVLN 263  
Db 350 IKONKEFFNQNTLVINKIFGVGKPKYKIDKADGKEGIWPKYKLPKIGTIEQYDPRKVS 409  
Qy 264 ASVQGSRRKRLRVYLHCTNTDNPYKEGDLT-----LYA-----INLHN 302  
Db 410 AYCDEKN---SIFISVDNTKSVTKAEYVTHNNLGGFWESCGEAYANESRSLINAFN 466  
Qy 303 VTKYLR-----PYPFSNKQVDKYLRLPLGPHGLLS 333  
Db 467 EGLHFNVSFKPSIFQDVRVKYLYNKYDGGFLS 500

RESULT 14  
F95146  
DNA topoisomerase I [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
R:Accession: F95146  
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzap  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: F95146  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-699 <KUR>  
A:Cross-references: UNIPROT:Q97QF3; UNIPARC:UPI0000003BFE; GB:AE005672; PIDN:AAK75367.1;  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SPI263  
C:Superfamily: DNA topoisomerase I

Query Match 4.7%; Score 95; DB 2; Length 699;  
Best Local Similarity 24.0%; Pred. No. 13;  
Matches 60; Conservative 39; Mismatches 75; Indels 76; Gaps 17;  
Qy 103 NAKLYGPDVGQPRRTAKMLKSLFKAG---GEVIDSVTH-HVYLN----- 144  
Db 67 NIRKGGLINDLKEAKKANKVFLASPDREGE---AISWHLAHLNLDENDANRVVNE 123  
Qy 145 -GRATATREDFLNPDVLDIFISSVQKVQFVWVSTRPGKVMWLGTSAYGGGAPLLSDTFA 203  
Db 124 ITKDAVKNAFKPRKIDMDLVDAQARRILDR-----VGYSIS-----PIL----- 165  
Qy 204 AGFWMLD-KLGLSARMGIEVVMQVFFGAGNYHLVD--EN-----FDPLPDYW-LSLLFKKL 256  
Db 166 ----WKKVKKGLSAGRVSQIALKLI-----IDRENEINAFQ-EEYWTVDVAFKK- 210  
Qy 257 VGTKVLMSVQGSRRKRLRVYLHCTNTDNPYKE- -GDLTLYAHLNHNVTKYLR-----L 309  
Db 211 -GKQFHAFSYGVDGKKMKL-----TSNNEVKVELSLRTSKDFSVQDVKKERKNAPL 263  
Qy 310 PYPFSNKQVD 319  
Db 264 PYTSSNQMD 273

RESULT 15  
D98014

DNA topoisomerase (EC 5.99.1.2) [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
R:Accession: D98014  
R:Authors: Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Bargett, S.; DeHoff, B.S.; Eberle, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Meyer, P.; Sun, P.M.; Winkler, M.E.  
J:Biochem Biophys Res Commun 193, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.; J.A. Bacteriol. 183, 5709-5717, 2001  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A57872; MUID:21429245; PMID:11544234  
A:Accession: D98014  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-701 <KUR>  
A:Cross-references: UNIPROT:Q8DP19; UNIPARC:UPI000000E35E2; GB:AE007317; PIDN:AAK99944.1;  
C:Genetics:  
A:Gene: topA  
C:Superfamily: DNA topoisomerase I  
C:Keywords: isomerase

Query Match	4.7%;	Score 95;	DB 2;	Length 701;
Best Local Similarity	24.0%;	Pred. No. 13;		
Matches	60;	Conservative	39;	Mismatches 75; Indels 76; Gaps 17;
Qy	103	NAKLYPDVQCPRRKTAKMLKSFKAG----	GEVIDSVTWH-HYYLN-----	144
Db	73	NIRKGGLINDLKEAKKANKVFLASDPDREGE--	AISWHLAHILNDENDANRVVFNE	129
Qy	145	-GRATATREDFLNPDVLDFITSSVQKVFQVVESTRPGKKVWLGETSSAVGGGAPLLSDTFA	203	
Db	130	ITKDAVKNAKPEPRKIDMDIVDAQARRIDRL-----	VGISIS-----	171
Qy	204	AGFMWLD-KLGLSARMGIEVVMRQVFFGAGNYHLVD--EN----	FDPLPDYW-LSLLFPKKL	256
Db	172	-----WKVKVKGLSAGRQSTALKLI-----	IDRENEINARQP-EEYVTVDVAFFK-	216
Qy	257	VGTKVLWASVOGSKRRLKRYLVLHCTNDNPRYKE--	GDLTLYAIINLHNVTKYL-	309
Db	217	-GTKQFHASPYGVDGKQMKL-----	TSNNEVKVELSRLTSKDPFVQVDKKERKNAPL	269
Qy	310	PYPFSNKKQVD	319	
Db	270	PYTTSSMQMD	279	

Search completed: June 5, 2006, 12:21:43  
Job time : 14.8514 secs